

# Outer Bounds on the Storage-Repair Bandwidth Tradeoff of Exact-Repair Regenerating Codes

Birenjith Sasidharan, N. Prakash, M. Nikhil Krishnan, Myna Vajha, Kaushik Senthoor,  
and P. Vijay Kumar

(email: biren@ece.iisc.ernet.in, prakashn@mit.edu, {nikhilmk,myna,kaushik.sr,vijay}@ece.iisc.ernet.in)

## Abstract

In this paper three outer bounds on the storage-repair bandwidth (S-RB) tradeoff of regenerating codes having parameter set  $\{(n, k, d), (\alpha, \beta)\}$  under the exact-repair (ER) setting are presented. The tradeoff under the functional-repair (FR) setting was settled in the seminal work of Dimakis et al. that introduced the framework of regenerating codes as well as a subsequent paper by Wu. While it is known that the ER tradeoff coincides with the FR tradeoff at the extreme points of the tradeoff, known respectively as the minimum-storage-regenerating (MSR) and minimum-bandwidth-regenerating (MBR) points, its characterization on the interior points remains open.

The first outer bound presented here termed as the *repair-matrix bound*, in conjunction with a recent code construction known as *improved layered codes* characterizes the normalized ER tradeoff for the case of  $(n, k = 3, d = n - 1)$ . The repair-matrix bound is derived by building on top of the techniques introduced by Shah et al. and applies to every parameter set  $(n, k, d)$ . It was earlier proved by Tian that the ER tradeoff lies strictly away from the FR tradeoff for the specific case  $(n = 4, k = 3, d = 3)$ . The repair-matrix bound shows that a non-vanishing gap exists between the ER and FR tradeoffs for *every* parameter set  $(n, k, d)$ .

The second outer bound builds upon a bound due to Mohajer and Tandon and improves the bound using the very same techniques introduced in the Mohajer-Tandon paper and for this reason, is termed here as the *improved Mohajer-Tandon bound*. While for  $d = k$  the improved Mohajer-Tandon bound performs on par with the Mohajer-Tandon bound, for  $d > k$  there is a significant improvement in the region of the tradeoff away from the MSR point. In the vicinity of the MSR point however, the repair-matrix bound outperforms the improved Mohajer-Tandon bound.

In the third and final result, we restrict our focus to linear codes, and present an outer bound for the normalized ER tradeoff applicable to linear codes for the case  $k = d$ . In conjunction with the well-known class of *layered codes*, our third outer bound characterizes the normalized ER tradeoff in the case of linear codes for the case  $k = d = n - 1$ . This bound is derived by analyzing the rank-structure of a parity-check matrix for a linear ER code.

## Index Terms

Distributed storage; regenerating codes; exact-repair; storage-repair-bandwidth tradeoff; tradeoff characterization; outer bounds.

## I. INTRODUCTION

### A. Regenerating Codes

In the regenerating-code framework [3], all symbols are drawn from a fixed finite field  $\mathbb{F}$  whose size is the power of a prime. The size of the field does not play an important role in the present paper and for this reason does not appear in our notation for the field. Data pertaining to a file comprised of  $B$  symbols is encoded into a set of  $n\alpha$  coded symbols and then stored across  $n$  nodes in the network with each node storing  $\alpha$  coded symbols.

This research is supported in part by the National Science Foundation under Grant 1421848 and in part by an India-Israel UGC-ISF joint research program grant. Birenjith Sasidharan would like to thank the support of TCS Research Scholar Programme Fellowship awarded to him. N. Prakash was a PhD student at IISc, Bangalore, and also an intern at NetApp, Bangalore during the duration of this work. M. Nikhil Krishnan and Myna Vajha would like to thank the support of Visvesvaraya PhD Scheme for Electronics & IT awarded by Department of Electronics and Information Technology, Government of India. A portion of the material in this paper was presented in part at the 2014 IEEE International Symposium on Information Theory [1], and in part at the 2015 IEEE International Symposium on Information Theory [2].

A data collector should be able to retrieve the file downloading entire data from any  $k$  nodes. Furthermore,  $k$  is the minimum such number that allows reconstruction of the file. In the event of a node failure<sup>1</sup>, node repair is accomplished by having the replacement node connect to any  $d$  nodes and download  $\beta \leq \alpha$  symbols from each node with  $\alpha \leq d\beta < B$ . These  $d$  nodes are referred to as helper nodes. From the minimality of  $k$ , it can be shown that  $d$  must lie in the range

$$k \leq d \leq n - 1.$$

The quantity  $d\beta$  is called as the repair bandwidth. Here one makes a distinction between functional and exact repair. By functional repair (FR), it is meant that a failed node will be replaced by a new node such that the resulting network continues to satisfy the data-collection and node-repair properties defining a regenerating code. An alternative to functional repair is *exact repair* (ER) under which one demands that the replacement node store precisely the same content as the failed node. From a practical perspective, ER is preferred at least for two reasons. Firstly, the algorithms pertaining to data collection and node repair remain static for the ER case. Secondly if the ER code is linear, then it permits the storage of data in systematic form, which facilitates operations under paradigms such as MapReduce [6]. We will use  $\mathcal{P}_f$  to denote the *full parameter set*  $\mathcal{P}_f = \{(n, k, d), (\alpha, \beta)\}$  of a regenerating code and use  $\mathcal{P}$  when we wish to refer to only the parameters  $(n, k, d)$ .

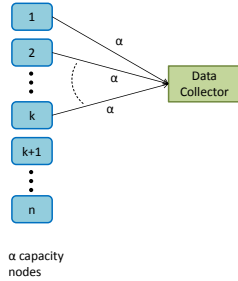


Fig. 1. Data collection.

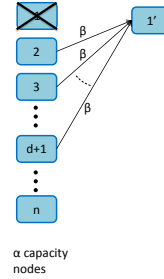


Fig. 2. Node repair.

### B. The Storage-Repair Bandwidth Tradeoff

A cut-set bound based on network-coding concepts, tells us that given a code parameter set  $\mathcal{P}_f$ , the maximum possible size  $B$  of a regenerating code is upper bounded as [3],

$$B \leq \sum_{\ell=0}^{k-1} \min\{\alpha, (d - \ell)\beta\}. \quad (1)$$

The derivation of the bound in (1) makes use of only FR constraints, and therefore it is valid for both FR and ER codes. An FR code  $\hat{\mathcal{C}}$  is said to be optimal if the file size  $\hat{B}$  of  $\hat{\mathcal{C}}$  achieves the cut-set bound in (1) with equality, and further, that if either  $\alpha$  or  $\beta$  is reduced, equality fails to hold in (1). The existence of such codes has been shown in [3], using network-coding arguments related to multicasting [7]. In general, we will use  $\hat{\mathcal{C}}$ ,  $\hat{B}$  etc to denote symbols relating to an optimal FR code while reserving  $\mathcal{C}$ ,  $B$  etc. to denote symbols relating to an ER code.

Given  $\mathcal{P}$  and  $B$ , there are multiple pairs  $(\alpha, \beta)$  that satisfy (1). It is desirable to minimize both  $\alpha$  as well as  $\beta$  since minimizing  $\alpha$  reduces storage requirements, while minimizing  $\beta$  results in a storage solution that minimizes repair bandwidth. It is not possible to minimize both  $\alpha$  and  $\beta$  simultaneously and thus there is a tradeoff between choices of the parameters  $\alpha$  and  $\beta$ . This tradeoff will be referred to as Storage-Repair Bandwidth (S-RB) tradeoff under functional repair. Since much of the emphasis of the current paper is upon the distinction between the S-RB tradeoffs under functional and exact repair, we will use FR tradeoff and ER tradeoff to refer respectively, to the two tradeoffs. The two extreme points in the FR tradeoff are termed the *minimum storage regeneration* (MSR) and

<sup>1</sup>Though regenerating codes are defined for the case of single node-failures, there are later works that looked into the case of simultaneous failure of multiple nodes, and studied cooperative repair in such a situation [4], [5]. However in this paper, we focus only on single node-failures.

*minimum bandwidth regeneration* (MBR) points respectively. The parameters  $\alpha$  and  $\beta$  for the MSR point on the tradeoff can be obtained by first minimizing  $\alpha$  and then minimizing  $\beta$  to yield

$$B = k\alpha, \quad \alpha = (d - k + 1)\beta. \quad (2)$$

Reversing the order leads to the MBR point which thus corresponds to

$$B = \left( dk - \binom{k}{2} \right) \beta, \quad \alpha = d\beta. \quad (3)$$

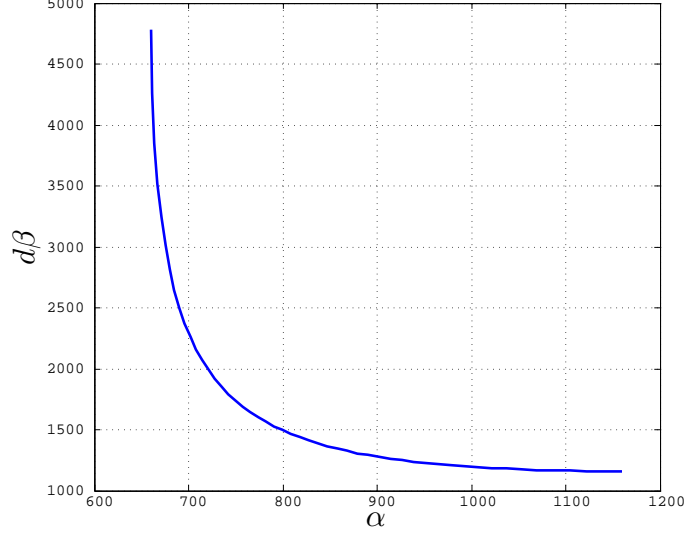


Fig. 3. FR Tradeoff. Here  $(n = 60, k = 51, d = 58, B = 33660)$ .

The remaining points on the tradeoff will be referred to as *interior points*. As the tradeoff is piecewise-linear, there are  $k$  points of slope discontinuity, corresponding to

$$\alpha = (d - \mu)\beta, \quad \mu \in \{0, \dots, k - 1\}.$$

Setting  $\mu = k - 1$  and  $0$  respectively, yields the MSR and MBR points. The remaining values of  $\mu \in \{1, \dots, k - 2\}$  correspond to interior points with slope-discontinuity. Interior points where there is no slope discontinuity can be specified by setting,

$$\begin{aligned} \alpha &= (d - \mu)\beta - \theta, \quad \theta \in [0, \beta) \\ &= (d - \mu)\beta - \nu\beta, \quad \nu \in [0, 1), \end{aligned} \quad (4)$$

with  $\mu \in \{0, 1, \dots, k - 2\}$ . When  $\mu = k - 1$ , we always set  $\nu = 0$ . We will refer to the pair  $(\alpha, \beta)$  as an *operating point* of the regenerating code. The tradeoff between  $\alpha$  and  $d\beta$  is plotted in Fig. 3 for  $(n = 131, k = 120, d = 130)$  and file size  $B = 725360$ .

The results in the present paper pertain to the ER tradeoff. Several ER code constructions [8]–[14] are now available that correspond to the MSR and the MBR points of the FR tradeoff. Thus the end points of the ER tradeoff coincide with those of the FR tradeoff. However, characterization of the interior points of the ER tradeoff remains an open problem in general.

### C. The Normalized ER Tradeoff and ER-Code Symmetry

For a given parameter set  $\mathcal{P} = (n, k, d)$ , there are several known constructions for an ER code, each of which is valid only for a restricted set of file sizes. Since the ER tradeoff for a fixed  $(n, k, d)$  varies with file size  $B$ , comparison across code constructions is difficult. For this reason, we normalize  $(\alpha, \beta)$  by the file size  $B$ . The tradeoff between  $\bar{\alpha} = \frac{\alpha}{B}$  and  $\bar{\beta} = \frac{\beta}{B}$  thus obtained for a fixed value of  $(n, k, d)$ , will be referred to here as the

*normalized ER tradeoff*. The tuple  $(\bar{\alpha}, \bar{\beta})$  is referred to as the *normalized operating point* of a regenerating code. Throughout the remainder of this paper, we will work only with the normalized version of the ER tradeoff.

Given a regenerating code  $\mathcal{C}$  associated to parameter set  $\mathcal{P}$  and file size  $B$ , the parameters of the code are clearly invariant to coordinate (i.e., node) permutation. Given an ER code  $\mathcal{C}$ , we can vertically stack the  $n!$  codewords obtained by encoding independent files using all possible node permutations of  $\mathcal{C}$ . The resultant stack of  $n!$  codewords may be regarded as a single new ER regenerating code  $\mathcal{C}'$  where the parameters  $(n, k, d)$  remain the same, but where the parameters  $(\alpha, \beta)$  and  $B$  are each scaled up multiplicatively, by a factor of  $n!$ . It is clear that  $\mathcal{C}'$  is symmetric in the sense that the amount of information contained in a subset  $A \subset [n]$  of nodes depends only upon the size  $|A|$  of  $A$ , and not upon the particular choice of nodes lying in  $A$ . This symmetry carries over even in the case of repair data transferred by a collection  $D$  of  $d = |D|$  nodes for the replacement of a fixed node. Such codes will be referred to as *symmetric* ER codes. Since the normalized values  $(\bar{\alpha}, \bar{\beta})$  of  $\mathcal{C}'$  remain the same as that of  $\mathcal{C}$ , there is no change in operating point on the normalized ER tradeoff in going from  $\mathcal{C}$  to  $\mathcal{C}'$ . Thus, given our focus on the normalized tradeoff, it is sufficient to consider *symmetric* ER codes. This observation was first made by Tian in [15].

#### D. Results

Though the complete characterization of normalized ER tradeoff for every parameter set remains an open problem, much progress has been made. It was shown in [13], that apart from the MBR point and a small region adjacent to the MSR point, there do not exist ER codes whose  $(\alpha, d\beta)$  values correspond to coordinates of an interior point on the FR tradeoff. However, the authors of [13] did not rule out the possibility of approaching the FR tradeoff asymptotically i.e., as the file size  $B \rightarrow \infty$ . It was first shown by Tian in [15] that the ER tradeoff lies strictly away from the FR tradeoff. This was accomplished by using an information theory inequality prover [16] to characterize the normalized ER tradeoff for the particular case of  $(n, k, d) = (4, 3, 3)$  and showing it to be distinct from the FR tradeoff. The results in the [15] were however, restricted to the particular case  $(n, k, d) = (4, 3, 3)$ .

That the ER tradeoff lies strictly above the FR tradeoff for *any* value of the parameter set  $(n, k, d)$  was first shown in [1]. The first result in the present paper is to show an outer bound on the normalized ER tradeoff for every parameter set  $(n, k, d)$ , and is stated in Thm. III.4. We refer to this outer bound as the *repair-matrix bound*. This outer bound in conjunction with a code construction appearing in [17], characterizes the normalized ER tradeoff for the parameter set  $(n, k, d)$  for  $k = 3$ ,  $d = n - 1$  and any  $n \geq 4$ .

Two outer bounds on the normalized ER tradeoff appeared subsequently in [18] and [19]. In [18], the author presents two bounds on the ER file size. In the first bound, he builds on top of the techniques presented in [15] and derives a bound that applies to a larger set of parameters. The second bound is obtained by taking a similar approach as in [1], and is shown to improve upon the one given in [1]. In [19], the author provides an upper bound on ER file size, that is non-explicit in general. However for the case of linear codes, the bound can be computed to obtain an explicit expression for any parameter set  $(n, k, d)$ . A second paper by Tian, [20], characterizes the ER tradeoff for  $(n = 5, k = 4, d = 4)$  with the help of a class of codes known as the *layered codes* introduced in [21]. A different approach adopted to derive an outer bound on the normalized ER tradeoff is presented in [22]. In [22], Mohajer et al. derived an outer bound for general  $(n, k, d)$  that turns out to be optimal for the special case of  $(n, k = n - 1, d = n - 1)$  in a limited region of  $\bar{\beta} \leq \frac{2\bar{\alpha}}{k}$  close to the MBR point. Optimality follows from the fact that a code construction due to Goparaju et al. in [23] meets their outer bound in the region  $\bar{\beta} \leq \frac{2\bar{\alpha}}{k}$ . We will refer to this outer bound in [22] as the *Mohajer-Tandon bound*.

The second result of the present paper is an improvement upon the Mohajer-Tandon bound for the case  $k < d$ . We make use of the very same techniques introduced in [22] to arrive at this improved bound. This bound is stated in Thm. V.1, and we refer to it as the *improved Mohajer-Tandon bound*. While the improved Mohajer-Tandon bound performs better whenever  $k < d$ , it coincides with the Mohajer-Tandon bound when  $k = d$ . The repair-matrix bound still performs better than the improved Mohajer-Tandon bound in a region close to the MSR point. The theorem below essentially combines the repair-matrix bound and the improved Mohajer-Tandon bound.

**Theorem I.1.** *Let*

$$B_1 = \sum_{i=0}^{k-1} \min\{\alpha, (d-i)\beta\} - \delta,$$

where  $\delta$  is as defined in (29), and it corresponds to the repair-matrix bound. Let  $B_2$  be the expression on the RHS in (43), corresponding to the improved Mohajer-Tandon bound. Then the ER file size  $B$  is bounded by,

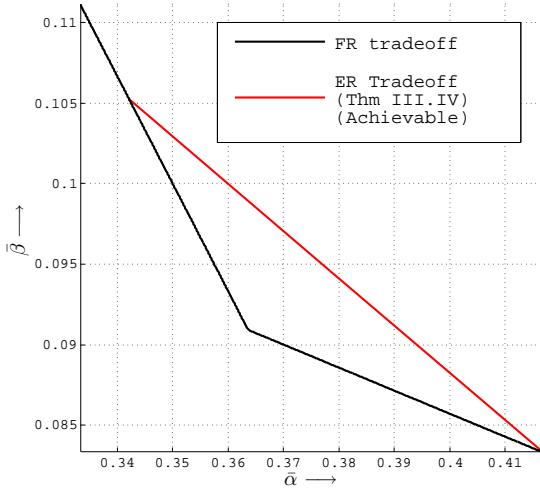
$$B \leq \min\{B_1, B_2\}.$$

The final result presented in this paper is under the restricted setting of linear codes. For the case of  $(n \geq 4, k = n - 1, d = n - 1)$ , we characterize the normalized ER tradeoff under this setting. This is done by deriving an explicit upper bound on the file size  $B$  of a ER linear regenerating code for the case  $k = d = n - 1, n \geq 4$ . The outer bound remains valid for the general case  $k = d$  even when  $d < n - 1$ . For the case of  $(n, k = n - 1, d = n - 1)$ , the outer bound matches with the region achieved by the layered codes. This result, which first appeared in [2], is stated below:

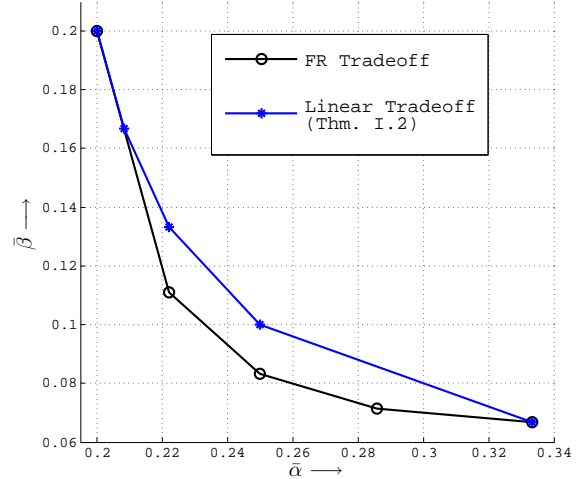
**Theorem I.2.** Consider an exact repair linear regenerating code, having parameters  $(n, k = n - 1, d = n - 1), (\alpha, \beta), n \geq 4$ . Then, the file size  $B$  of the code is upper bounded by

$$B \leq \begin{cases} \left\lfloor \frac{r(r-1)n\alpha + n(n-1)\beta}{r^2 + r} \right\rfloor, & \frac{d\beta}{r} \leq \alpha \leq \frac{d\beta}{r-1}, & 2 \leq r \leq n-2 \\ (n-2)\alpha + \beta, & \frac{d\beta}{n-1} \leq \alpha \leq \frac{d\beta}{n-2} \end{cases}. \quad (5)$$

We remark that there are no known instances of non-linear codes that violate the above outer bound derived under the linear setting. In an independent work [24], the authors also derive the normalized linear ER tradeoff for the case  $(n, k = n - 1, d = n - 1)$ , but the tradeoff is expressed in an implicit manner as the solution to an optimization problem.



(a) For  $k = 3, d = n - 1$ , codes in [17] achieves our repair-matrix bound. The example here is  $(n = 6, k = 3, d = 5)$ .



(b) For  $k = d = n - 1$ , our outer bound matches the achievable region of layered codes, thus characterizing the tradeoff under linear setting. The example here is  $(n = 6, k = 5, d = 5)$ .

Fig. 4. Characterization of normalized ER Tradeoff.

In Fig. 4, we plot the cases in which our outer bounds characterize the normalized ER tradeoff. In Fig. 5, we do a performance comparison of various known bounds.

### E. Our Approach

The present paper derives outer bounds on the normalized ER tradeoff of a regenerating code with full-parameter-set  $\mathcal{P}_f = \{(n, k, d), (\alpha, \beta)\}$ . Since every ER code is an FR code, it is clear that the normalized ER tradeoff lies on or above and to the right, of the normalized FR tradeoff in the  $(\bar{\alpha}, \bar{\beta})$ -plane. When we say that the normalized ER tradeoff *lies above* the normalized FR tradeoff, we imply that for given  $(n, k, d)$  there is at least one value of normalized parameter  $\bar{\beta}_0$  such that the corresponding normalized values  $\bar{\alpha}_{\text{ER}}$  and  $\bar{\alpha}_{\text{FR}}$  satisfy  $\bar{\alpha}_{\text{ER}} > \bar{\alpha}_{\text{FR}}$ . An equivalent definition in terms of the file size  $B$  is given as follows. For given  $(n, k, d)$ , let  $\hat{B}_0 := \hat{B}_{\text{opt}}(\alpha_0, \beta_0)$

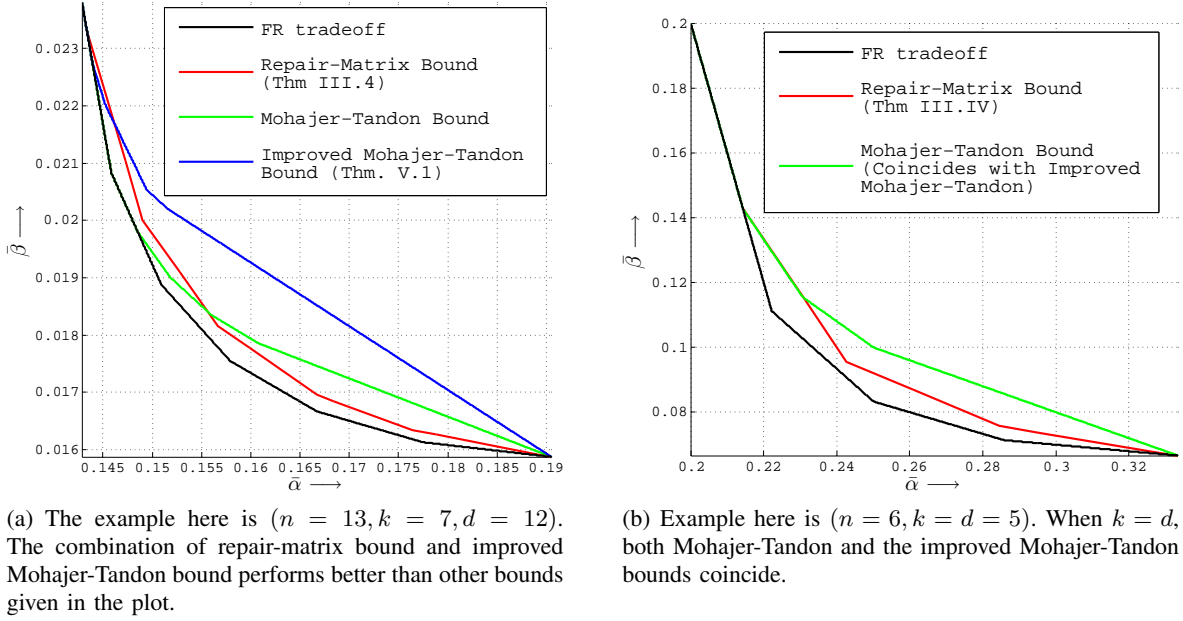


Fig. 5. Performance comparison of various outer bounds.

denote the optimal FR file size at an operating point  $(\alpha_0, \beta_0)$  with  $\alpha_0 = (d - \mu)\beta_0 - \nu\beta_0$  as in (4). Thus  $(\frac{\alpha_0}{\beta_0}, \frac{\beta_0}{\beta_0})$  is a point lying on the normalized FR tradeoff. Suppose that the maximum file size of an ER code as a function of  $(\alpha, \beta)$  is

$$B(\alpha, \beta) = \hat{B}(\alpha, \beta) - \epsilon(\alpha, \beta)$$

for some non-negative function  $\epsilon(\alpha, \beta)$ . Let  $\epsilon_0 = \epsilon(\alpha_0, \beta_0)$ . Then the normalized operating points  $(\bar{\alpha}_{\text{ER}}, \bar{\beta}_{\text{ER}})$  for an optimal ER code as given by

$$\begin{aligned} \bar{\beta}_{\text{ER}} &= \frac{\beta_0}{B(\alpha_0, \beta_0)} = \frac{1}{\frac{\hat{B}_0}{\beta_0} - \frac{\epsilon_0}{\beta_0}} \\ \bar{\alpha}_{\text{ER}} &= \frac{\alpha_0}{B(\alpha_0, \beta_0)} = \frac{1}{\frac{\hat{B}_0}{\alpha_0} - \frac{\epsilon_0}{\alpha_0}} = \frac{1}{\frac{\hat{B}_0}{\alpha_0} - \frac{\epsilon_0}{\beta_0} \frac{1}{(d-\mu-\nu)}} \end{aligned}$$

will be bounded away from  $(\frac{\alpha_0}{\beta_0}, \frac{\beta_0}{\beta_0})$  if  $(\frac{\epsilon_0}{\beta_0})$  does not vanish to zero. It follows that an upper bound on the file size  $B$  of an ER code

$$B \leq B_{\text{upper}}(\alpha, \beta),$$

such that

$$\lim_{\beta \rightarrow \infty} \frac{\hat{B}(\alpha, \beta) - B_{\text{upper}}(\alpha, \beta)}{\beta} > 0 \quad (6)$$

for some  $(\mu, \nu)$  will equivalently define a bound on the normalized ER tradeoff that lie strictly above the normalized FR tradeoff. Throughout the paper, our approach therefore will be to derive upper bounds on ER file size that satisfy the criterion in (6).

If the full parameter set of a regenerating code has  $n > (d + 1)$ , then by restricting attention to a set of  $(d + 1)$  nodes, one obtains a regenerating code with  $n = (d + 1)$  with all other parameters remaining unchanged. It follows from this that any upper bound on the size  $B$  corresponding to full parameter set  $\{(n = (d + 1), k, d), (\alpha, \beta)\}$  continues to hold for the case  $n > (d + 1)$  with the remaining parameters left unchanged. Keeping this in mind, we will assume throughout that  $n = (d + 1)$ .

A key technique used in the paper is to lower bound the difference  $\epsilon = \hat{B}_{\text{opt}}(\alpha, \beta) - B(\alpha, \beta)$  between the file size of an optimal FR code and an ER code. The total information content in a regenerating code can be accumulated

from a set  $\{1, 2, \dots, k\}$  of  $k$  nodes. The conditional entropy of the  $(i+1)$ -th node data conditioned on the data accumulated from previous  $i$ ,  $0 \leq i \leq k-1$  nodes is compared against the corresponding value of an optimal FR code, and the difference is defined to be  $\omega_i$ . It follows that  $\epsilon$  is the sum of all  $\{\omega_i\}_{i=0}^{k-1}$ . Our approach is to relate  $\{\omega_i\}_{i=0}^{k-1}$  in terms of entropy of certain collections of repair data, and eventually find an estimate on  $\epsilon$ . Along the way, we construct a *repair matrix* as an arrangement of random variables corresponding to repair data in a  $((d+1) \times (d+1))$ -sized matrix. Many properties pertaining to the inherent symmetry of regenerating code become clear from the repair-matrix perspective, and we use it as a tool in our proofs.

A different approach is used in deriving an upper bound on the ER file size of a linear regenerating code. Here we focus on a parity-check matrix  $H$  of a linear ER code, and construct an augmented parity-check matrix  $H_{\text{repair}}$  of size  $(n\alpha \times n\alpha)$  that captures the exact-repair properties. A block-matrix structure is associated to  $H_{\text{repair}}$ , and thereby we identify  $n$  thick columns  $\{H_1, H_2, \dots, H_n\}$  of  $H_{\text{repair}}$  with  $H_i$  associated to the node  $i$ . Here we mean by a thick column a collection of  $\alpha$  columns. Let us denote by  $\delta_i$  the incremental rank added by  $H_i$  to the collection of  $(i-1)\alpha$  vectors in  $\{H_j \mid 1 \leq j < i\}$ . We estimate lower bounds on  $\{\delta_i\}_{i=1}^n$  that will eventually lead to a lower bound on the rank of  $H$ . It is clear that the file size  $B$  is the dimension of the code, and therefore a lower bound on the rank of  $H$  results in an upper bound on the file size.

## F. Organization of the Paper

In Sec. II, we describe the result of Shah et al. showing the non-existence of ER codes operating on the FR tradeoff. In Sec. III-C, we present an upper bound on the ER file size. In Sec. IV, we review the various upper bounds on ER file size that are known in the literature. In Sec. V, we develop on the existing Mohajer-Tandon bound, and make an improvement upon that to get a better bound when  $d > k$ . In Sections VI, VII, VIII, we focus on upper bounds on file size under linear setting. We characterize the normalized ER tradeoff for the case  $(n, k = n-1, d = n-1)$  in Sec. VIII, while the proof techniques are illustrated for a particular case of  $(n = 5, k = 4, d = 4)$  in Sec. VII. In Sec. IX, we discuss the achievability of the outer bounds on normalized ER tradeoff derived at earlier sections.

## II. THE NON-EXISTENCE OF ER CODES ACHIEVING FR TRADEOFF

As mentioned in Sec. I-D, it was shown in [13] that apart from the MBR point and a small region adjacent to the MSR point, there do not exist ER codes whose  $(\alpha, d\beta)$  values correspond to coordinates of an interior point on the FR tradeoff. The theorem in [13] due to Shah et al. is stated below.

**Theorem II.1.** (Theorem 7 in [13]) *For any given values of  $(n, k \geq 3, d)$ , ER codes do not exist for the parameters  $(\alpha, \beta, B)$  lying at an interior point on the FR tradeoff except possibly for the case*

$$(d-k+1)\beta \leq \alpha \leq \left[ (d-k+2) - \frac{d-k+1}{d-k+2} \right] \beta. \quad (7)$$

The region

$$\{(\alpha, \beta) \mid (d-k+1)\beta \leq \alpha \leq \left[ (d-k+2) - \frac{d-k+1}{d-k+2} \right] \beta\}$$

on which the theorem does not claim the non-existence of ER codes is referred to as the *near-MSR region*. The Theorem II.1 however did not rule out the possibility of approaching the FR tradeoff asymptotically i.e., as the file size  $B \rightarrow \infty$ . As mentioned earlier, this question was answered by Tian in the negative in [15] for the specific case when  $(n, k, d) = (4, 3, 3)$ .

In this section, we will describe the approach taken by Shah et al. in proving Theorem II.1 in terms of the notation to be used in the present paper. We begin with some notation and definitions. Let  $\mathcal{C}$  be an ER regenerating code over  $\mathbb{F}$  having file size  $B$  and full-parameter set  $\mathcal{P}_f = \{(n, k, d), (\alpha, \beta)\}$ . We regard the message symbols as a collection of  $B$  random variables taking on values in  $\mathbb{F}$  and use  $M$  to denote the  $(1 \times B)$  random vector whose components are the  $B$  message symbols. We use  $p_M(\cdot)$  to denote the joint probability distribution of the  $M$  random variables. All other random variables pertaining to the regenerating code are functions of the components of  $M$ , and satisfy probability distributions that are induced by  $p_M$ .

We will use  $[i]$ ,  $1 \leq i \leq n$  to denote the set  $\{1, 2, \dots, i\}$  and define  $[0]$  to be the empty set  $\phi$ . For  $1 \leq i \leq j \leq n$ , we use  $[i:j]$  to denote the set  $\{i, i+1, \dots, j\}$ . Whenever we write  $[i:j]$  with  $i > j$ , it will be assumed to be the empty set. On occasion, we will run into a set of random variables of the form  $W_A$  where  $A$  is the empty set,  $W_A$  should again be interpreted as the empty set.

### A. The Repair Matrix and the Constraints Imposed By Exact-Repair

As made clear in Sec. I-E, we assume that  $n = d + 1$  without loss of generality. Let  $W_x, 1 \leq x \leq n$  denote the random variable corresponding to the contents of a node  $x$ . Given a subset  $A \subseteq [n]$ , we use

$$W_A = \{W_x \mid x \in A\}$$

to denote the contents of nodes indexed by  $A$ . Clearly,

$$H(W_x) \leq \alpha. \quad (8)$$

Let  $S_x^y, x, y \in [n], x \neq y$  denote the random variables corresponding to the helper data sent by the helper node  $x$  to the replacement of a failed node  $y$ . This is well defined because under the assumption  $n = (d + 1)$ , there is just one set of  $d$  helper nodes for any failed node. Given a pair of subsets  $X, Y \subseteq [n]$ , we define  $S_X^Y = \{S_x^y \mid x \in X, y \in Y, x \neq y\}$ . We use the short-hand notation  $S_X$  to indicate  $S_X^X$ . From the definition of regenerating codes, it follows that

$$H(S_x^y) \leq \beta. \quad (9)$$

In (8, 9), information is measured in units of  $\log_2(|\mathbb{F}|)$  bits. The collection of random variables  $\{S_x^y \mid x \in [d+1], y \in [d+1], x \neq y\}$  can schematically be represented using a  $(d + 1) \times (d + 1)$  matrix  $\mathcal{S}$  with empty cells along the diagonal as shown in Fig. 6(a). The rows in this matrix correspond to the helper nodes and the columns to nodes undergoing repair. The  $(x, y)$ th entry of this matrix, thus corresponds to  $S_x^y$ . We will refer to  $\mathcal{S}$  as the *repair matrix*. The subset of  $\mathcal{R}$  appearing below the diagonal and above the diagonal are denoted by  $\mathcal{R}_L$  and  $\mathcal{R}_U$  respectively.

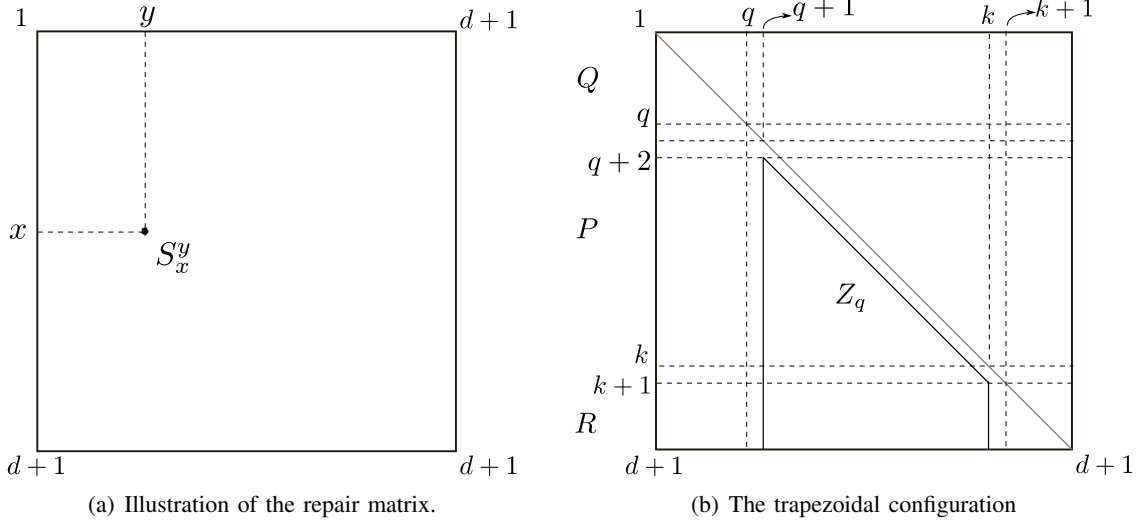


Fig. 6. The repair matrix and the trapezoidal configuration

Apart from the constraints given in (8), (9), the requirements of data reconstruction and exact-repair impose further constraints. The constraint due to data reconstruction is given by either of the following two equivalent statements:

$$H(W_A) = B, |A| \geq k, \quad (10)$$

$$H(M \mid W_A) = 0, |A| \geq k. \quad (11)$$

For every  $i \in [n]$ , the exact-repair condition imposes the constraint

$$H(W_i \mid S_{\mathcal{D}}^i) = 0, |\mathcal{D}| = d, i \notin \mathcal{D}. \quad (12)$$



### B. Trapezoidal Configurations in the Repair Matrix

Throughout the discussion taking place in Sections up to III, we will assume that there is a fixed numbering of the  $n = (d + 1)$  nodes in the network. In (10), the file size  $B$  is expressed as the joint entropy of a collection  $k$  random variables  $\{W_1, W_2, \dots, W_k\}$ . It is possible to express  $B$  as the joint entropy of other subsets of random variables, in particular those involved in node repair. An example, important for the discussion to follow, appears below. Let  $q$  be an integer lying in the range  $0 \leq q \leq k$  and set

$$\begin{aligned} Q &= \{1, 2, \dots, q\} \\ P &= \{q + 1, q + 2, \dots, k\} \\ R &= \{k + 1, k + 2, \dots, (d + 1)\}. \end{aligned}$$

Note that  $Q, P, R$  are all functions of the integer  $q$ . When  $q = 0$ , we will set  $Q$  to be the empty set  $\phi$ . Note that  $P = [k] \setminus Q$  and  $R = [k + 1 : d + 1]$ . We define:

$$Z_q = \mathcal{R}_L \cap S_{[d+1]}^P \quad (13)$$

$$X_q = \mathcal{R}_L \cap S_P. \quad (14)$$

Then we can write  $B$  as:

$$\begin{aligned} B &= H(W_Q, W_P) \\ &= H(W_Q, W_P, Z_q) \\ &= H(W_Q, Z_q) + H(W_P | W_Q, Z_q) \\ &= H(W_Q, Z_q) \end{aligned}$$

where (15) follows from the exact-repair condition (12). The collection  $Z_q$  of random variables forms a trapezoidal region within the repair matrix as shown in Fig.6(b). We refer to  $(W_Q, Z_q)$ ,  $q \in \{0, 1, \dots, k\}$  as a *trapezoidal configuration*. The set  $Z_q$  is said to be the *trapezoid* corresponding to the trapezoidal configuration  $(W_Q, Z_q)$ . It is clear that  $Z_q = X_q \uplus S_R^P$ . Next we proceed to define a *sub-trapezoid* of the trapezoid  $Z_q$ . Let  $T = \{q + 1, q + 2, \dots, q + t\} \subseteq P$  be a subset of size  $0 \leq t \leq k - q$  of  $P$ . Then we define the subset  $Z_{q,t}$  of  $Z_q$  as:

$$Z_{q,t} := \mathcal{R}_L \cap S_{[d+1]}^T.$$

The set  $Z_{q,t}$  also forms a trapezoidal region in  $\mathcal{R}$  and is called a sub-trapezoid of the trapezoid  $Z_q$ . Here again, we define  $X_{q,t}$  as:

$$X_{q,t} := S_T \cap Z_{q,t},$$

and it follows that  $Z_{q,t} = X_{q,t} \uplus S_{R \cup (P \setminus T)}^T$ . A sub-trapezoid is illustrated in Fig. 7.

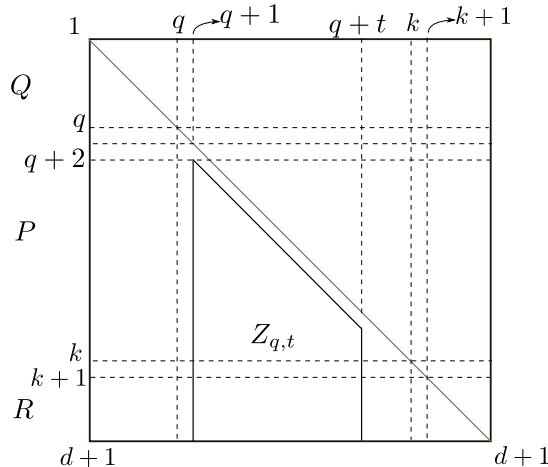


Fig. 7. Illustration of the sub-trapezoid  $Z_{q,t}$ .

For every trapezoidal configuration  $(W_Q, Z_q)$  indexed by  $q = 0, 1, \dots, k$ , we have the identity

$$B = H(W_Q, Z_q), \quad (15)$$

and the corresponding inequality obtained by repeatedly applying the union bound  $H(X_1, X_2) \leq H(X_1) + H(X_2)$ , i.e.,

$$\begin{aligned} B &\leq H(W_Q) + H(Z_q) \\ &\leq H(W_Q) + H(X_q) + H(S_R^P) \end{aligned} \quad (16)$$

$$\leq q\alpha + \binom{k-q}{2}\beta + (d+1-k)(k-q)\beta. \quad (17)$$

We define for  $q \in \{0, 1, 2, \dots, k\}$ , the quantities:

$$B_q := q\alpha + \binom{k-q}{2}\beta + (d+1-k)(k-q)\beta.$$

### C. The Argument For Non-existence

Let us consider an ER code operating at the point  $(\alpha, \beta)$  satisfying  $\alpha = (d - \mu)\beta$ . For this value of  $\alpha$ , as shown below, the FR bound gives us  $B_{\mu+1}$  as the upper bound on file size:

$$\begin{aligned} B &\leq \sum_{i=0}^{k-1} \min\{\alpha, (d-i)\beta\} \\ &= (\mu+1)\alpha + \sum_{i=\mu+1}^{k-1} (d-i)\beta \\ &= (\mu+1)\alpha + \sum_{j=0}^{k-\mu-2} (d-k+1+j)\beta \\ &= (\mu+1)\alpha + (d-k+1)(k-\mu-1)\beta + \binom{k-\mu-1}{2}\beta \\ &= B_{\mu+1}. \end{aligned}$$

Thus if an ER code is optimal with respect to the FR tradeoff at the point  $\alpha = (d - \mu)\beta$ , from equations (15) and (16), with  $q = (\mu + 1)$ , one obtains that such a code must satisfy:

$$H(Z_{\mu+1} | W_{[\mu]}) = H(Z_{\mu+1}) = \binom{k-\mu-1}{2}\beta + (d+1-k)(k-\mu-1)\beta, \quad (18)$$

i.e., the union bound on  $Z_{\mu+1}$  must hold with equality. That means that all the random variables in  $Z_{\mu+1}$  are mutually independent. However, it is shown by Shah et al. in [13] that this is not possible if an ER code lies at an interior point except for the near-MSR region and the MBR point. To prove this result, the authors of [13] focus on a subset  $S_m^L$  of the repair matrix where  $m \in [n]$  and  $L \subseteq [n]$  are arbitrarily chosen from  $[n]$  while satisfying the conditions  $|L| := \ell < k$  and  $m \notin L$ . The subset  $S_m^L$  is of course, the union of helper data sent by a single node  $m$  to the nodes in  $L$ . We can write

$$\begin{aligned} H(S_m^L) &= H(S_m^L | W_L) + I(S_m^L : W_L) \\ &\leq H(S_m^L | W_L) + I(W_m : W_L). \end{aligned} \quad (19)$$

It can be shown that (see [13])

$$H(S_m^L | W_L) = 0, \quad \ell \geq \mu + 1, \quad (20)$$

and that

$$I(W_m : W_L) = \beta, \quad \ell = \mu + 1. \quad (21)$$

As a consequence, we have that

$$H(S_m^L) = \beta, \ell = \mu + 1. \quad (22)$$

It follows that

$$H(S_m^J) \leq \beta, \text{ for any } J \subseteq [n] \text{ with } |J| < \mu + 1.$$

In particular this is true if  $J$  is of size  $|J| = 2$ . On the other hand, optimality with respect to the FR bound assumes that each row in the trapezoidal region  $Z_q$  has joint entropy equal to the number of repair random variables  $S_x^y \in Z_q$  belonging to the row, times  $\beta$ . The bottom row of the trapezoid has  $(k - \mu - 1)$  entries and thus we clearly have a contradiction whenever  $(k - \mu - 1) \geq 2$ . The argument does not go through when  $(k - \mu - 1) \leq 1$ , i.e., when  $\mu \geq k - 2$ . This necessary condition on  $\mu$  underlies the fact that the non-existence of ER codes do not hold good in the near-MSR region. The proof given here is for the case when  $\alpha = (d - \mu)\beta$  is a multiple of  $\beta$ . This proof can be extended to the general case  $\alpha = (d - \mu)\beta - \theta$ , for  $0 < \theta < \beta$  as well. In the next section, we will exploit this contradiction to derive an upper bound on the file size of an ER code.

### III. AN UPPER BOUND ON THE ER FILE SIZE

In this section, we show that for *any* value of the parameter set  $(n, k, d)$ , the ER tradeoff lies strictly above the FR tradeoff, a result that was first established in [1]. As explained in Sec. I-E, we do this by deriving a tighter bound on file size  $B$  in the case of ER than is true under FR.

As mentioned in Sec. II-C, our approach to bounding the file size  $B$  is based on deriving estimates for the joint entropy of subsets of the repair matrix. First, we assume the existence of an ER code having parameters  $(n, k, d), (\alpha, \beta)$  whose file size  $B$  is of the form  $B = \hat{B} - \epsilon$  for some  $\epsilon \geq 0$ , where  $\hat{B}$  is the file size of an optimal FR code having the same parameter set  $\mathcal{P}$ . Next, we proceed to estimate the joint entropy of the subset  $Z_q$  corresponding to a trapezoidal configuration  $(W_Q, Z_q)$ . We estimate the joint entropy in two different ways and show that the two estimates are in contradiction unless the value of  $\epsilon$  lies above a threshold value  $\epsilon_{\min}$ . This allows us to replace  $B - \epsilon_{\min}$  as the revised bound on the file size under ER. We will also show that  $\epsilon_{\min}$  does not vanish as  $\beta \rightarrow \infty$ .

#### A. Preliminaries

Consider an optimal FR code  $\hat{\mathcal{C}}$  possessing the same set of parameters  $\mathcal{P}$  as the ER code  $\mathcal{C}$ . In what follows, given any deterministic or random entity associated with  $\mathcal{C}$ , we will use a hat to denote the corresponding entity in  $\hat{\mathcal{C}}$ . For example,  $\hat{B}$  denotes the file size of  $\hat{\mathcal{C}}$ . With this, we can write

$$\begin{aligned} \sum_{i=0}^{k-1} \min\{\alpha, (d-i)\beta\} &= \hat{B} = H(\hat{W}_{[k]}) \\ &= \sum_{i=0}^{k-1} H(\hat{W}_{i+1} \mid \hat{W}_{[i]}) \\ &\leq \sum_{i=0}^{k-1} \min\{\alpha, (d-i)\beta\}. \end{aligned}$$

It follows that in an optimal FR code  $\hat{\mathcal{C}}$ , we must have

$$H(\hat{W}_{i+1} \mid \hat{W}_{[i]}) = \min\{\alpha, (d-i)\beta\}, \quad 0 \leq i \leq (k-1).$$

Next, for  $0 \leq i \leq k-1$ , let us set:

$$\begin{aligned} \gamma_i &= \min\{\alpha, (d-i)\beta\}, \\ \omega_i &= \gamma_i - H(W_{i+1} \mid W_{[i]}), \end{aligned}$$

where  $\omega_i$  measures the drop in the conditional entropy  $H(W_{i+1} | W_{[i]})$  of an ER code in comparison with its value  $H(\hat{W}_{i+1} | \hat{W}_{[i]})$  in the case of an optimal FR code. A plot of  $\gamma_i$  as a function of  $i$  for a given operating point  $(\alpha, \beta)$  with  $\alpha = (d - \mu)\beta - \theta$ , appears in Fig. 8. We also note the following identities:

$$\epsilon = \sum_{i=0}^{k-1} \omega_i, \quad (23)$$

$$H(W_B | W_A) = \sum_{i=a}^{a+b-1} (\gamma_i - \omega_i), \quad (24)$$

where  $A = [a]$  and  $B = [a+1 \ a+b]$  and  $0 \leq a \leq a+b \leq k$ . The lemma below follows from these identities.

**Lemma III.1.** *Let  $(Q, Z_q)$  be a trapezoidal configuration for some  $q \in \{0, 1, \dots, k\}$ , and let  $Z_{q,t} \subseteq Z_q$  be a sub-trapezoid with  $0 \leq t \leq k - q$ . Then*

$$H(Z_{q,t} | W_Q) \geq \sum_{i=q}^{q+t-1} (\gamma_i - \omega_i)$$

*Proof:* By the exact-repair condition,  $H(Z_{q,t} | W_Q)$  is at least  $H(W_{[q+1 \ q+t]} | W_Q)$  and the result follows from (24).  $\blacksquare$

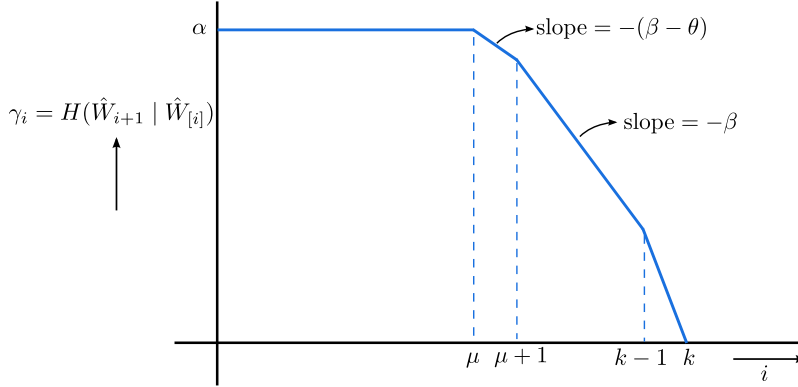


Fig. 8. The function  $\gamma_i$  versus  $i$  for  $\alpha = (d - \mu)\beta - \theta$ .

### B. Upper Bounds On Joint Conditional Entropies Of Repair Data

Let  $Q = [q]$ , and  $M, L$ , be two mutually disjoint subsets of  $[d+1] \setminus Q$  with  $\ell := |L|$ , and  $m := |M|$ . Then we can write

$$H(S_M^L | W_Q) = H(S_M^L | W_V, W_Q) + I(S_M^L : W_V | W_Q), \quad (25)$$

where in we take  $V \supset L$  as a superset of  $L$  with  $V \cap M = \emptyset$  and  $v := |V|$ . Our next objective is to estimate  $H(S_M^L | W_V, W_Q)$  and  $I(S_M^L : W_V | W_Q)$  in order to obtain an upper bound on  $H(S_M^L | W_Q)$ .

**Lemma III.2.** *Suppose  $\alpha = (d - \mu)\beta - \theta$  with  $\mu \in \{0, 1, \dots, k-1\}$  and  $\theta \in [0, \beta)$  except when  $\mu = k-1$ . Then for  $2 \leq \ell \leq v < k - q$ ,*

$$H(S_M^L | W_V, W_Q) \leq \begin{cases} \ell\theta + \ell\omega_{v-1+q}, & v = \mu + 1 - q \\ \ell\omega_{v-1+q}, & v > \mu + 1 - q. \end{cases}$$

*Proof:* Let  $\ell_0 \in L$ , and by symmetry  $H(S_M^{\ell_0} | W_V, W_Q)$  is same for every  $\ell_0 \in L$ . Define  $\tilde{V} = V \setminus \{\ell_0\}$ . Then we have

$$\begin{aligned} H(S_M^L | W_V, W_Q) &\leq \ell H(S_M^{\ell_0} | W_V, W_Q) \\ &= \ell \{H(S_M^{\ell_0}, W_{\ell_0} | W_{\tilde{V}}, W_Q) - H(W_{\ell_0} | W_{\tilde{V}}, W_Q)\} \\ &= \ell \{H(S_M^{\ell_0} | W_{\tilde{V}}, W_Q) + H(W_{\ell_0} | S_M^{\ell_0}, W_{\tilde{V}}, W_Q) - H(W_{\ell_0} | W_{\tilde{V}}, W_Q)\} \end{aligned}$$

By substituting bounds, we obtain for the case  $v - 1 + q > \mu$

$$\begin{aligned} H(S_M^L | W_V, W_Q) &\leq \ell\{m\beta + (d - v + 1 - q - m)\beta - (d - v + 1 - q)\beta + \omega_{v-1+q}\} \\ &= \ell\omega_{v-1+q}, \end{aligned}$$

and for the case  $v - 1 + q = \mu$ ,

$$\begin{aligned} H(S_M^L | W_L, W_Q) &\leq \ell\{m\beta + (d - v + 1 - q - m)\beta - (d - v + 1 - q)\beta + \theta + \omega_{v-1+q}\} \\ &= \ell\theta + \ell\omega_{v-1+q}. \end{aligned}$$

■

We remark here that in [18] the quantity  $H(S_M^L)$  is considered for obtaining a bound on ER file size. Our approach here is different in the sense that we estimate  $H(S_M^L)$  in terms of  $\{\omega_i\}_{i=0}^{k-1}$ . The second term in (25) can also be easily estimated in terms of  $\{\gamma_i, \omega_i\}_{i=0}^{k-1}$ :

$$\begin{aligned} I(S_M^L : W_V | W_Q) &\leq I(W_M : W_V | W_Q) \\ &= H(W_M | W_Q) - H(W_M | W_{Q \cup V}) \\ &= \left[ \sum_{i=q}^{q+m-1} (\gamma_i - \omega_i) \right] - \left[ \sum_{i=q+v}^{q+v+m-1} (\gamma_i - \omega_i) \right]. \end{aligned} \quad (26)$$

The Lemma III.2 along with (26) allows us to bound  $H(S_M^L | W_Q)$  from above given an operating point  $\alpha = (d - \mu)\beta - \theta$ . Calculations for the particular case of  $q = 0, m = 1$  taking values for  $v$  in  $\{\mu + 1, \mu + 2\}$  result in the following corollary.

**Corollary III.3.** *Let  $\alpha = (d - \mu)\beta - \theta$ . Then for  $m \notin L$  and  $\ell = |L|$ , we have*

$$H(S_m^L) \leq \beta + (\ell - 1)\theta + (\ell - 1)\omega_\mu + (\omega_\mu + \omega_{\mu+1}), \quad 2 \leq \ell \leq \mu + 1 \quad (27)$$

$$H(S_m^L) \leq 2\beta - \theta + (\ell - 1)\omega_{\mu+1} + (\omega_{\mu+1} + \omega_{\mu+2}), \quad 2 \leq \ell \leq \mu + 2. \quad (28)$$

### C. The Bound On ER File Size

In this section, we make use of Lem. III.1 and Cor. III.3 to derive an upper bound on the file size  $B$  of an ER code. This will also translate to an outer bound for the ER tradeoff.

**Theorem III.4.** *Let  $B$  denote the file size of a ER regenerating code with full-parameter set  $\mathcal{P}_f = \{(n, k, d), (\alpha, \beta)\}$ . Let  $\alpha = (d - \mu)\beta - \theta$ . Then the ER file size  $B$  is upper bounded by:*

1) For  $\mu = 0, 0 < \theta < \beta$ ,

$$B \leq \hat{B} - \epsilon_1$$

2) For  $\mu \in \{1, 2, \dots, k - 3\}, 0 \leq \theta < \beta$ ,

$$B \leq \hat{B} - \max\{\epsilon_0, \epsilon_1\}$$

3) For  $\mu = k - 2, 0 \leq \theta < \left(\frac{d-k+1}{d-k+2}\right)\beta$ ,

$$B \leq \hat{B} - \epsilon_0,$$

where  $\epsilon_0$  and  $\epsilon_1$  are as given in Tab. I.

*Proof:* The proof is relegated to the Appendix. ■

**Corollary III.5.** *When  $k \geq 3$ , the normalized ER tradeoff is strictly away from the normalized FR tradeoff for all normalized operating points  $(\bar{\alpha}, \bar{\beta})$  with  $\bar{\alpha} = (d - \mu)\bar{\beta} - \nu\bar{\beta}$  such that  $(\mu, \nu)$  falls in the range  $(\mu = 0, 0 < \nu < 1)$ ,  $(\mu \in \{1, 2, \dots, k - 3\}, 0 \leq \nu < 1)$  or  $(\mu = k - 2, 0 \leq \nu < \frac{d-k+1}{d-k+2})$ .*

Regime of $(\mu, \theta)$	Lower bounds $\epsilon_0$ , $\epsilon_1$ on $\epsilon = \hat{B} - B$
$\mu \in \{1, 2, \dots, k-2\}$ for all $\theta$ For $\mu = k-2$ , $\theta < \frac{d-k+1}{d-k+2}\beta$	Let $r_0 = \left\lfloor \frac{k-\mu}{\mu+1} \right\rfloor$ $\epsilon_0 = \begin{cases} \frac{(d-k+1)(k-\mu-1)(\beta-\theta) - \theta}{(d-k+1)(k-\mu) + 1}, & k - \mu < \mu + 1. \\ \frac{\left(d - \frac{(\mu+1)(r_0+3)}{2} + 2\right)r_0\mu(\beta-\theta) - \theta}{\left(d - \frac{(\mu+1)(r_0+3)}{2} + 2\right)r_0(\mu+1) + 1}, & k - \mu \geq \mu + 1. \end{cases}$
$\mu \in \{0, 1, \dots, k-3\}$ for all $\theta$ For $\mu = 0$ , $\theta \neq 0$	Let $r_1 = \left\lfloor \frac{k-\mu-1}{\mu+2} \right\rfloor$ $\epsilon_1 = \begin{cases} \frac{(d-k+1)[(k-\mu-3)\beta + \theta]}{(d-k+1)(k-\mu-1) + 1}, & k - \mu - 1 < \mu + 2. \\ \frac{\left(d - \frac{(\mu+2)(r_1+3)}{2} + 2\right)r_1[\mu\beta + \theta]}{\left(d - \frac{(\mu+2)(r_1+3)}{2} + 2\right)r_1(\mu+2) + 1}, & k - \mu - 1 \geq \mu + 2. \end{cases}$

TABLE I  
LOWER BOUNDS ON THE QUANTITY  $\hat{B} - B$

*Proof:* We will show that the upper bound on the file size given in III.4 satisfies the criterion in (6). Let

$$\delta = \begin{cases} \epsilon_1 & \mu = 0, \theta \neq 0 \\ \max\{\epsilon_0, \epsilon_1\} & \mu \in \{1, 2, \dots, k-3\} \\ \epsilon_0 & \mu = k-2, \theta < \frac{d-k+1}{d-k+2}\beta \end{cases} \quad (29)$$

Let  $\alpha$  be related to  $\beta$  as  $\alpha = (d-\mu)\beta - \theta = (d-\mu)\beta - \nu \cdot \beta$ ,  $\nu \in [0, 1)$  by a fixed pair  $(\mu, \nu)$  that falls in the range given. Then for a code with the file size  $B$ ,

$$\begin{aligned} \frac{\beta}{B} &\geq \frac{\beta}{\hat{B} - \delta}, \quad (\text{using Thm. III.4}) \\ &= \frac{\beta}{\hat{B}} \cdot \frac{1}{1 - \left(\frac{\delta}{\hat{B}}\right)} \\ &= \frac{\beta}{\hat{B}} \cdot \frac{1}{1 - \left(\frac{\delta}{\beta \sum_{i=0}^{k-1} \min\{(d-\mu)-\nu, (d-i)\}}\right)} \\ &\geq \frac{\beta}{\hat{B}} + \delta_0, \end{aligned}$$

for some  $\delta_0 > 0$ , determined by the constants  $\frac{\epsilon_0}{\beta}$  and  $\frac{\epsilon_1}{\beta}$ . It can be seen that  $\frac{\epsilon_0}{\beta}$  and  $\frac{\epsilon_1}{\beta}$  are independent of  $\beta, B$  and dependent only on the fixed values of  $\mu, \nu, k$  and  $d$ . This completes the proof. ■

#### IV. DISCUSSION ON VARIOUS KNOWN UPPER BOUNDS ON ER FILE SIZE

In this section, we briefly review the results from [15], [20], [18], [19], [22], all of them involving upper bounds on the ER file size. While bounds provided in [18], [19] are not explicit, those presented in [15], [20], [22] have got the form of explicit algebraic expressions.

### A. Review of the Bounds in [15], [20]

In [15], Tian characterized the optimal ER file size for the case of  $(n, k, d) = (4, 3, 3)$ . This was the first result establishing a non-vanishing gap for ER file size in comparison with the optimal FR file size. For the case of  $(n, k, d) = (4, 3, 3)$ , there are four bounds

$$B \leq B_q, \quad q = 0, 1, 2, 3, \quad (30)$$

that follow from considering all possible trapezoidal configurations. For a given operating point  $\alpha = (d - \mu)\beta - \theta$ , one of these bounds dominate over the others. By suitably modifying the information theory inequality prover software (see [16], [25]), Tian was able to characterize a bound

$$3B \leq 4\alpha + 6\beta,$$

that is different from (30). Recently in [20], Tian made further progress with his computational approach to provide an upper bound on the ER file size for  $(n, k, d) = (5, 4, 4)$ . In both the case of  $(4, 3, 3)$  and  $(5, 4, 4)$ , the bounds are achieved using the well-known class of layered codes [21]. These results are made part of the online collection of “Solutions of Computed Information Theoretic Limits (SCITL)” hosted at [26].

### B. Review of the Bound in [18]

In the second of two bounds presented in [18], Duursma considers the region  $Z_q$  in a trapezoidal configuration  $(Q, Z_q)$ , and tiles the region using rectangular blocks corresponding to random variables  $S_M^L$ , with  $m := |M|$ ,  $\ell := |L|$ . This approach is an extension of the tiling-with-line-segments method, introduced in [1] and used in the present paper in the derivation of Thm. III.4. Duursma extends the upper bound given in [1] to obtain a bound on  $H(S_M^L)$ , involving entropy expressions having a negative coefficient. Various carefully-chosen alternative bounds on  $B$  are used to cancel out these negative terms leading to the improved bound:

$$B + \sum_{(M,L) \in \mathcal{M}} \ell B \leq B_q + \sum_{(M,L) \in \mathcal{M}} (B_{r+m-1} + (\ell - 1)(B_{r+m-2} - \beta)), \quad (31)$$

where  $m := |M|$ ,  $\ell := |L|$  and  $r \geq \ell$  for every choice of  $(M, L)$ . In (31),  $\mathcal{M}$  denotes a set of possible tilings of the trapezoidal region  $Z_q$  using rectangular blocks, and  $B_q$  remains as defined in Section II-B. To obtain the best possible explicit bound, one would then proceed to minimize this expression over all possible tilings. It can easily be checked that the bound in (31) is tighter than the one given in (17), by a difference of at most  $\beta$ .

### C. Review of the Bound in [19]

In [19], Duursma augments the set of node random variables  $\{W_i\}_{i=1}^k$  with another set of random variables  $W'_{k+u}$  for  $1 \leq u \leq \nu$  satisfying

$$H(S_i^j | W'_{k+u}) \leq H(S_i^j | W_{[i+1,k]} W'_{[k+1,k+u-1]}) \text{ for } 1 \leq i < j \leq p, \quad (32)$$

for a given value of  $p$ ,  $0 \leq p \leq k$ . The bound on file size  $B$  is obtained as

$$(\nu + 1)B \leq (\nu + 1)B_{k-p} + \sum_{u=1}^{\nu} \left( H(W'_{k+u}) - \binom{p}{2} \beta \right),$$

where  $B_{k-p}$  is as defined earlier. This results in general, in an implicit bound as it is not clear how the random variables  $\{W'_{k+u}\}_{u=1}^{\nu}$  can be constructed. However, restricting to linear codes, the author is able to construct the  $\{W'_{k+u}\}$  resulting in an explicit bound for every parameter set  $(n, k, d)$ . This bound matches with the one proved in [2] for the special case of  $(k+1, k, k)$ -linear ER codes.

#### D. Review of the Bound in [22]

In this section, we give a complete description<sup>2</sup> of the proof of the bound due to Mohajer et al. in [22]. We start with recalling the bound given in (15) for a trapezoidal configuration  $(Q, Z_q)$ ,

$$\begin{aligned} B &\leq H(W_Q) + H(Z_q | W_Q) \\ &= H(W_Q) + H(X_q, S_R^P | W_Q), \end{aligned} \quad (33)$$

where the sets  $P$ ,  $Q$ , and  $R$  are as defined in Sec. II-B. For convenience of notation, we modify the indexing of elements in sets  $P$ ,  $Q$  and  $R$ , without making any change in their respective sizes. Thus the sets  $Q, P, R$  are defined by the same value of  $q$ , and hence the bound in (15) remains unaltered. With respect to the modified indexing,  $Q = \{-1, -2, \dots, -q\}$ ,  $P = \{1, 2, \dots, p := k - q\}$  and  $R = \{k + 1, k + 2, \dots, d + 1\}$ . Continuing from (33), we write

$$\begin{aligned} B &\leq H(W_Q) + H(X_q, S_R^P | W_Q) \\ &\leq q\alpha + \underbrace{\sum_{i=1}^p H(S_i^{[i-1]} | W_Q)}_{\mathcal{R}(p)} + H(S_R^P | W_Q). \end{aligned} \quad (34)$$

Instead of invoking the union bound as done in (16), the entropic term  $\mathcal{R}(p) := \sum_{i=1}^p H(S_i^{[i-1]} | W_Q)$  is canceled out with the help of other expressions for  $B$ . In (35) that follows, the authors over-count conditional node entropy  $H(W_i | W_{[i-1]})$  as  $\alpha$ , and later subtract out the error introduced in doing so. This leads to a different expression for  $B$ :

$$\begin{aligned} B &= H(W_Q) + \sum_{i=1}^p H(W_i | W_Q) - \sum_{i=1}^p I(W_i; W_{[i-1]} | W_Q) \\ &\leq q\alpha + p\alpha - \sum_{i=1}^p I(S_i^{[i-1]}; S_{[i-1]}^i | W_Q) \\ &= \underbrace{k\alpha - \sum_{i=1}^p H(S_i^{[i-1]} | W_Q)}_{\mathcal{R}(p)} - \underbrace{\sum_{i=1}^p H(S_{[i-1]}^i | W_Q)}_{\mathcal{C}(p)} + \underbrace{\sum_{i=1}^p H(S_{[i-1]}^i, S_i^{[i-1]} | W_Q)}_{\mathcal{J}(p)}. \end{aligned} \quad (35)$$

While (35) allows cancellation of  $\mathcal{R}(p)$  in (34), it introduces new entropic terms  $\mathcal{C}(p)$  and  $\mathcal{J}(p)$ . A third expression for  $B$  is obtained by over-counting entropy of columns in the trapezoidal region  $Z_q$  using union bound, and then subtracting out the error introduced in doing so.

$$\begin{aligned} B &\leq H(W_Q, S_{[d+1]}^P) \\ &\leq q\alpha + \sum_{i=1}^p H(S_{[d+1]}^i | W_Q) - \sum_{i=1}^p I(S_{[d+1]}^i; S_{[d+1]}^{[i-1]} | W_Q) \\ &\leq q\alpha + \sum_{i=1}^p H(S_{[i-1]}^i | W_Q) + \sum_{i=1}^p H(S_{[i+1] \dots [d+1]}^i | W_Q) - \sum_{i=1}^p I(S_{[d+1]}^i; S_{[d+1]}^{[i-1]} | W_Q). \end{aligned} \quad (36)$$

The following straightforward lemma is useful in producing a lower bound for  $I(S_{[d+1]}^i; S_{[d+1]}^{[i-1]} | W_Q)$ .

**Lemma IV.1.** *Let  $X, Y, Z, U$  be random variables such that  $Z = f_1(X, U) = f_2(Y, U)$  for some deterministic functions  $f_1, f_2$ . Then*

$$I(X : Y | U) \geq H(Z | U).$$

<sup>2</sup>We have simplified the proof to some extent, and therefore certain arguments differ from what is presented in [22].



By invoking Lem. IV.1 along with identifying  $Z = \{S_{[i-1]}^i, S_i^{[i-1]}\}$ ,  $X = S_{[d+1]}^i$ ,  $Y = S_{[d+1]}^{[i-1]}$  and  $U = W_Q$ , it follows that

$$I\left(S_{[d+1]}^i; S_{[d+1]}^{[i-1]} | W_Q\right) \geq H\left(S_{[i-1]}^i, S_i^{[i-1]} | W_Q\right), \quad (37)$$

and substituting (37) back in (36), the authors obtain the bound

$$B \leq q\alpha + \underbrace{\sum_{i=1}^p H\left(S_{[i-1]}^i | W_Q\right)}_{\mathcal{C}(p)} + \sum_{i=1}^p H\left(S_{[i+1 \ d+1]}^i | W_Q\right) - \underbrace{\sum_{i=1}^p H\left(S_{[i-1]}^i, S_i^{[i-1]} | W_Q\right)}_{\mathcal{J}(p)}. \quad (38)$$

Summation of (34) (35) and (38) eliminates  $\mathcal{R}(p)$ ,  $\mathcal{C}(p)$  and  $\mathcal{J}(p)$ , and results in the bound:

$$3B \leq (3k - 2p)\alpha + \sum_{i=1}^p H\left(S_{[i+1 \ d+1]}^i | W_Q\right) + H\left(S_R^P | W_Q\right). \quad (39)$$

By applying union bound, it follows that

$$B \leq \min_{0 \leq p \leq k} \frac{(3k - 2p)\alpha + \frac{p(2(d-k)+p+1)\beta}{2} + (d - k + 1) \min\{\alpha, p\beta\}}{3}. \quad (40)$$

To our knowledge, the bound in (40) due to Mohajer et al. remains the best known upper bound on ER file size in the region away from the MSR point.

## V. AN IMPROVED UPPER BOUND ON ER FILE SIZE

In this section, we first propose an improvement over the bound in [22], that is described in Sec. IV-D. The authors of [22] apply union bound on the last two terms in (39) to obtain the final bound. But it is possible to avoid the union bound for the term  $H\left(S_R^P | W_Q\right)$  when  $d \gg k$ .

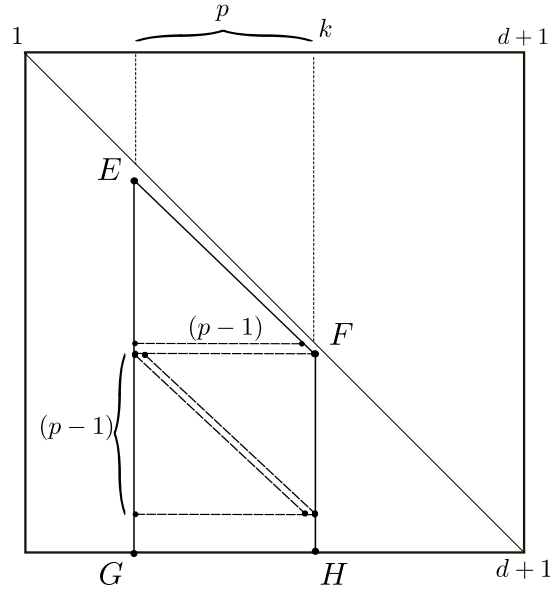


Fig. 9. The splitting up of the region corresponding to  $S_R^P$ . In this example,  $a = 1$ ,  $b > 0$ .

The Fig. 9 illustrates the region  $S_R^P$  as it is viewed on the repair matrix. The rectangular region  $S_R^P$ , denoted by  $\Gamma$ , is of width  $p$  and height  $(d - k + 1)$ . Let us write

$$d - k + 1 = a(p - 1) + b, \quad 0 \leq b < (p - 1).$$

Then  $\Gamma$  can be split into  $(a + 1)$  sub-rectangles  $\Gamma_1, \Gamma_2, \dots, \Gamma_{a+1}$  of equal width  $p$ , and  $\Gamma_i, 1 \leq i \leq a$  have the same height  $(p - 1)$ . The last sub-rectangle  $\Gamma_{a+1}$  is of height  $b$ , and it vanishes in the case  $b = 0$ . Each rectangle

$\Gamma_i, 1 \leq i \leq a$  is further split into two isosceles right triangles  $\Gamma_{i1}, \Gamma_{i2}$  of base  $(p-1)$  as illustrated in Fig. 9. By symmetry, we can write

$$\begin{aligned} H(S_P^R|W_Q) &\leq aH(\Gamma_1|W_Q) + H(\Gamma_{a+1}|W_Q) \\ &\leq 2aH(\Gamma_{11}|W_Q) + H(\Gamma_{a+1}|W_Q) \\ &\leq 2a \sum_{i=1}^p H(S_i^{[i-1]}|W_Q) + b \min\{\alpha, p\beta\}. \end{aligned} \quad (41)$$

We improve upon the the bound in (34) by substituting (41), and obtain that

$$B \leq q\alpha + (1+2a) \sum_{i=1}^p H(S_i^{[i-1]}|W_Q) + b \min\{\alpha, p\beta\}. \quad (42)$$

This modification only affects the coefficient of the term  $\mathcal{R}(p)$ . The cancellation of the term  $\mathcal{R}(p)$  is possible by appropriately scaling the bounds in (35) and (38). This results in an improved bound whenever  $a \geq 1$ , and is stated in the theorem below. We refer to this bound as the *improved Mohajer-Tandon bound*.

**Theorem V.1.** *The ER file size  $B$  of regenerating code with full-parameter set  $\mathcal{P}_f = \{(n, k, d), (\alpha, \beta)\}$  is bounded by*

$$B \leq \min_{0 \leq p \leq k} \frac{\alpha(2(k-p)(1+a) + k(1+2a)) + b \min\{\alpha, p\beta\} + \frac{(1+2a)p(2(d-k)+p+1)\beta}{2}}{3+4a}, \quad (43)$$

where  $d - k + 1 = a(p-1) + b$  and  $0 \leq b < (p-1)$ .

We remark that the improved Mohajer-Tandon bound relies upon the same techniques introduced by Mohajer et al. of coming up with various expressions for  $B$  allowing one to cancel out entropic terms that are otherwise difficult to estimate. Our incremental contribution is limited to identifying the symmetry in certain entropic terms as seen in the pictorial depiction on a repair matrix, and leveraging upon this symmetry to avoid certain union bounds. When  $d > k$ , the bound in Thm. V.1 leads to an outer bound on normalized ER tradeoff, that lies above the one due to (40). A principal result of the paper stated in Thm. I.1 follows by combining both the Thm. III.4 and the Thm. V.1.

## VI. A DUAL-CODE-BASED APPROACH TO BOUNDING THE ER FILE SIZE FOR LINEAR CODES

In this section, we investigate the maximum possible ER file size under the restricted setting of linear regenerating codes. Let  $\mathcal{C}_{\text{lin}}$  denote a linear ER code with full-parameter set  $\mathcal{P}_f = \{(n, k, d), (\alpha, \beta)\}$ . We will continue to use  $B$  to denote the file size. By linear, we mean that (a) the encoding mapping that converts the  $B$  message symbols to  $n\alpha$  coded symbols is linear, (b) the mapping that converts the node data into repair data that is transmitted during the repair of a failed node is linear and furthermore, (c) the mappings that are involved during data collection from a set of  $k$  nodes and regeneration of a failed node using repair data from a set of  $d$  nodes are linear. A linear regenerating code can be viewed as a linear block-code with length  $n\alpha$  over  $\mathbb{F}$  such that every set of  $\alpha$  symbols (taken in order without loss of generality) are bunched together to correspond to a node.

### A. The Parity-Check Matrix And Its Properties

Since  $\mathcal{C}_{\text{lin}}$  is a linear code, we can associate a generator matrix to the code. Let  $G$  of size  $(B \times n\alpha)$  denote a generator matrix of  $\mathcal{C}_{\text{lin}}$ . Without loss of generality, we assume that the first  $\alpha$  columns of  $G$  generate the contents of the first node, the second  $\alpha$  columns of  $G$  generate the contents of the second node, and so on. The first  $\alpha$  columns taken together will be referred to as the first thick column of  $G$ . Similarly, the second thick column consists of columns from  $\alpha + 1$  to  $2\alpha$ , and so on. Overall, we will have  $n$  thick columns in  $G$ . Let  $H$  denote a parity-check matrix having size  $(n\alpha - B) \times n\alpha$ . The row-space of  $H$  is the dual code of  $\mathcal{C}_{\text{lin}}$ . The definition of thick columns directly carries over to  $H$ . For any set  $S \subseteq [n]$ , we write  $H|_S$  to denote the restriction of  $H$  to the thick columns indexed by the set  $S$ . From definitions, we have that

$$B = \text{rank}(G) = n\alpha - \text{rank}(H). \quad (44)$$

By (44), it is sufficient to obtain a lower bound on  $\text{rank}(H)$  to bound  $B$  from above. This is precisely the approach taken here.

In the following two lemmas, we will translate the properties of data collection and exact-repair as properties of the parity-check matrix  $H$ . We remark here that these observations are already made in [18].

**Lemma VI.1** (Data Collection). *Let  $H$  be a parity-check matrix of an ER linear regenerating code. Then  $\text{rank}(H|_S) = (n - k)\alpha$ , for any  $S \subseteq [n]$  such that  $|S| = n - k$ .*

*Proof:* This is a re-statement of Part (1) of Proposition 2.1 of [18], and is equivalent to the data collection property. ■

**Lemma VI.2** (Exact Repair). *Assume that  $d = n - 1$ . Then the row space of  $H$  of an ER linear regenerating code contains a collection of  $n\alpha$  vectors that can be arranged as the rows of an  $(n\alpha \times n\alpha)$  matrix  $H_{\text{repair}}$ , which can be written in the block-matrix form:*

$$H_{\text{repair}} = \begin{bmatrix} A_{1,1} & A_{1,2} & & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ & & \ddots & \\ A_{n,1} & A_{n,2} & & A_{n,n} \end{bmatrix}, \quad (45)$$

where  $A_{i,i}$  is defined to be the identity matrix  $I_\alpha$  of size  $\alpha$  and  $A_{i,j}$  denotes an  $\alpha \times \alpha$  matrix such that  $\text{rank}(A_{i,j}) \leq \beta$ ,  $1 \leq i, j \leq n$ ,  $i \neq j$ .

*Proof:* The first  $\alpha$  rows of the form

$$[ I_\alpha \mid A_{1,2} \mid \cdots \mid A_{1,n} ]$$

can be obtained by the parity-check equations that are necessitated by the exact-repair requirement of the first node. In a similar manner, there must be parity-check equations that must allow repair of every node. These parity-check equations can be arranged to obtain the matrix  $H_{\text{repair}}$ . The requirements on the ranks of the sub-matrices  $A_{ij}$  follow from the definition of regenerating codes, and the fact that  $d = n - 1$ . In fact, the proof is indicated in Part (2) of Proposition 2.1 of [18]. ■

For the case of  $d = k = n - 1$ , the matrix  $H_{\text{repair}}$  as given in Lem. VI.2 satisfies the condition given in Lem. VI.1, and therefore  $H_{\text{repair}}$  by itself defines an  $(n, k = n - 1, d = n - 1)(\alpha, \beta)$  regenerating code. Since  $\text{rank}(H) \geq \text{rank}(H_{\text{repair}})$ , and our interest lies in regenerating codes having maximal file size, we will assume that  $H = H_{\text{repair}}$  while deriving a lower bound on  $\text{rank}(H)$  for the case of  $d = k = n - 1$ .

### B. A Proof Of FR Bound For ER Linear Codes Using Dual Code

In this section, we will present a simple proof of the FR bound (1) for ER linear regenerating codes. Our proof of Theorem I.2 will be built up on the proof of (1) that is presented here.

As earlier, let  $\mathcal{C}_{\text{in}}$  denote an  $(n, k, d = n - 1)(\alpha, \beta)$  linear regenerating code, and let the matrix  $H$  generate the dual code of  $\mathcal{C}$ . The key idea of the proof is to obtain a lower bound on the column rank of the matrix  $H$ . We use the notation  $\rho(\cdot)$  to denote the rank of a matrix. Let us define the quantities  $\delta_j$ ,  $1 \leq j \leq n$  as follows:

$$\delta_1 = \rho(H|_{[1]}), \quad (46)$$

$$\delta_j = \rho(H|_{[j]}) - \rho(H|_{[j-1]}), \quad 2 \leq j \leq n. \quad (47)$$

Next, we make the following claims:

$$\delta_j = \rho(A_{j,j}) \quad (48)$$

$$= \alpha, \quad 1 \leq j \leq n - k \quad (49)$$

$$\delta_j \geq (\alpha - (j - 1)\beta)^+, \quad n - k + 1 \leq j \leq n. \quad (50)$$

Here we have set  $a^+ := \max(a, 0)$ . The first claim (48) follows from the fact that any  $n - k$  thick columns of  $H$  has rank given by  $(n - k)\alpha$  as required by Lem. VI.1. To show the second claim (50), one needs to first focus on the  $j^{\text{th}}$  thick row of  $H_{\text{repair}}$ . By  $j^{\text{th}}$  thick row, we mean the set of rows starting from  $(j - 1)\alpha + 1$  and reaching up to  $j\alpha$  of  $H_{\text{repair}}$ . Next observe that

$$\delta_j \geq \left( \rho(A_{j,j}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}) \right)^+ \quad (51)$$

$$\begin{aligned} &= \left( \rho(I_\alpha) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}) \right)^+ \\ &\geq (\alpha - (j - 1)\beta)^+, \quad n - k + 1 \leq j \leq n, \end{aligned} \quad (52)$$

where (52) holds true since  $\rho(A_{i,j}) \leq \beta$  by Lem. VI.2. Thus we have shown (50). Next, invoking (48) and (52), we bound the column-rank of  $H$  from below as:

$$\text{rank}(H) = \sum_{j=1}^n \delta_j \quad (53)$$

$$\geq (n - k)\alpha + \sum_{j=n-k+1}^n (\alpha - (j - 1)\beta)^+. \quad (54)$$

An illustration of arriving at (51) and (54) is given in Fig. 10. Consequently, it follows that


$$\rho(H) \geq \rho(A_{1,1}) + \sum_{j=2}^5 \left( \rho(A_{j,j}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}) \right)^+$$


Fig. 10. A lower bound on  $\rho(H)$ , for the case of  $(n = 5, k = 4, d = 4)$ . Each term indexed by  $j$  in the summation correspond to a lower bound on the incremental rank  $\delta_j$ . This bound is obtained by looking at the sub-matrices in  $j^{\text{th}}$  thick row.

$$B = n\alpha - \rho(H) \quad (55)$$

$$\leq n\alpha - (n - k)\alpha - \sum_{j=n-k+1}^n (\alpha - (j - 1)\beta)^+ \quad (56)$$

$$= \sum_{j=0}^{k-1} \min(\alpha, (n - 1 - j)\beta). \quad (57)$$

For  $d < n - 1$ , the proof follows by first puncturing the code on any  $(n - d - 1)$  nodes to form a  $(n' = d + 1, k, d)$  ER linear regenerating code and then invoking the above analysis on the resultant new code. The way we express incremental ranks  $\{\delta_j\}$  in (48) and (51) will turn out to be useful in deriving a strong upper bound on the file size of linear ER codes in Sections VII and VIII.

## VII. AN UPPER BOUND ON THE FILE SIZE OF LINEAR ER CODES FOR THE CASE $(n = 5, k = 4, d = 4)$

In this section, we obtain a new upper bound on the file size of a linear ER code for parameters  $(n = 5, k = 4, d = 4)$ . Taken along with the achievability using layered codes (see Sec. IX-B), we characterize the tradeoff for this case. As mentioned earlier, our technique is to lower bound the rank of the parity-check matrix  $H$ , leading to an upper bound on the file size by (44). The lower bound on  $\rho(H)$  that we derive here is in general tighter than what is obtained in (54). The principal result of this section is stated in Thm. VII.1 below. Most of the ideas that are developed in the proof of Thm. VII.1 will later be used in the next section to prove a general result for the case of  $(n, k = n - 1, d = n - 1)$ .

**Theorem VII.1.** *Consider an ER linear regenerating code  $\mathcal{C}_{lin}$  with full-parameter set  $\{(n = 5, k = 4, d = 4), (\alpha, \beta)\}$ . Let  $H$  denote a parity-check matrix of  $\mathcal{C}_{lin}$ . Then*

$$\rho(H) \geq \begin{cases} \left\lceil \frac{10(\alpha - \beta)}{3} \right\rceil, & 2\beta \leq \alpha \leq 4\beta \\ \left\lceil \frac{15\alpha - 10\beta}{6} \right\rceil, & \frac{4}{3}\beta \leq \alpha \leq 2\beta \\ 2\alpha - \beta, & \beta \leq \alpha \leq \frac{4}{3}\beta \end{cases} \quad (58)$$

Note that  $\alpha = \beta$  and  $\alpha = 4\beta$  correspond to the MSR and MBR points respectively for the case of  $(n = 5, k = 4, d = 4)$ . Next, we observe that for a fixed  $\beta$ , the bound given in (58) corresponds to a piecewise linear curve with  $\alpha$  on the  $X$ -axis and  $\rho(H)$  on the  $Y$ -axis. Non-linear ceiling operation  $\lceil \cdot \rceil$  is used in (58) to enforce integrality requirements on  $\rho(H)$ . However, it may be removed considering that  $\rho(H)$  always takes integer values. We can view (58) as a combination of the following three inequalities without paying attention to the limited range of  $\alpha$ :

$$\rho(H) \geq \frac{10(\alpha - \beta)}{3} \quad (59)$$

$$\rho(H) \geq \frac{15\alpha - 10\beta}{6} \quad (60)$$

$$\rho(H) \geq 2\alpha - \beta. \quad (61)$$

Here (61) follows from (54) since  $\alpha \geq \beta$  and  $(\alpha - (j - 1)\beta)^+ \geq 0$  for  $3 \leq j \leq 5$ . Therefore, we need to prove only the remaining two inequalities (59) and (60) to complete the proof of Thm. VII.1. We proceed to prove them by obtaining two lower bounds to the incremental thick-column-rank of  $H$  that are stronger than what is given in (52). To make this point clear upfront, a comparison of the bounds in (54) and (58) is shown in Fig. 11.

### A. Proof of Theorem VII.1

We begin with setting up some notation. For any matrix  $B$  over  $\mathbb{F}$ , we denote by  $\mathcal{S}(B)$  the column space of  $B$ . Note that  $\rho(B)$  is the same as the dimension of the vector space  $\mathcal{S}(B)$ . Next, we define  $H^{(5)} = H_{repair}$ , where  $H_{repair}$  is as defined in (45). Let the matrix  $H_j^{(5)}$  denote the  $j^{\text{th}}$  thick column of  $H^{(5)}$ ,  $1 \leq j \leq 5$ , i.e.,  $H^{(5)} = [H_1^{(5)} H_2^{(5)} H_3^{(5)} H_4^{(5)} H_5^{(5)}]$ . Next, we define matrices  $H_j^{(4)}$ ,  $2 \leq j \leq 5$  such that the columns of  $H_j^{(4)}$  form a basis for the vector space  $\mathcal{S}(H_j^{(5)}) \cap \mathcal{S}(H^{(5)}|_{[j-1]})$ . Next, we define  $H^{(4)}$  as

$$H^{(4)} = [H_2^{(4)} H_3^{(4)} H_4^{(4)} H_5^{(4)}].$$

For convenience of notation, we have used  $H_2^{(4)}$  to denote the first thick column of  $H^{(4)}$ . Similarly,  $H^{(3)}$  is obtained from  $H^{(4)}$ , where columns of  $H_j^{(3)}$  form a basis for  $\mathcal{S}(H_j^{(4)}) \cap \mathcal{S}(H^{(4)}|_{\{2, \dots, j-1\}})$ :

$$H^{(3)} = [H_3^{(3)} H_4^{(3)} H_5^{(3)}].$$

Let  $A_{i,j}^{(\ell)}$  denote the  $i^{\text{th}}$  thick row of  $H_j^{(\ell)}$ . An illustration of the block-matrix representations of  $H^{(5)}, H^{(4)}$  and  $H^{(3)}$  is given in Fig. 12.

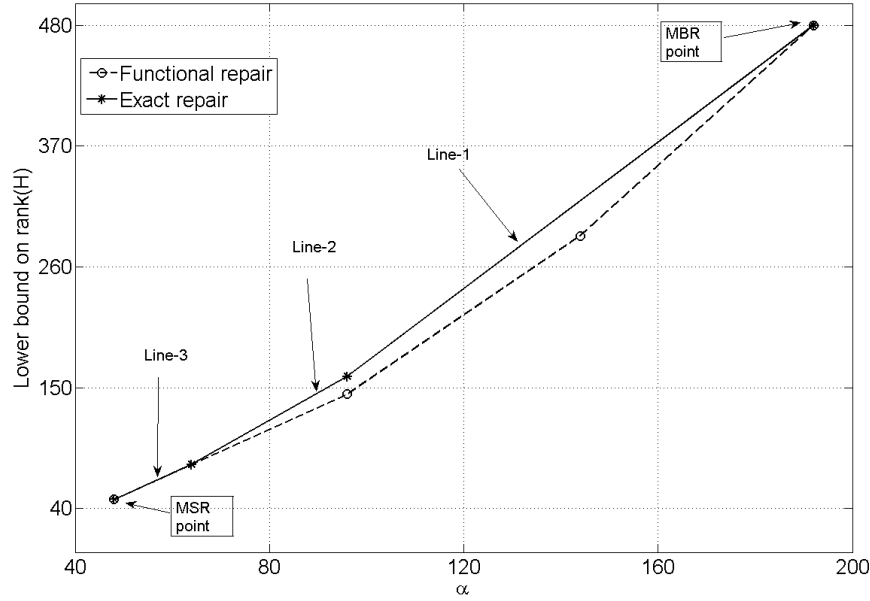


Fig. 11. Comparison of the lower bounds on  $\rho(H)$  as function of  $\alpha$ , for the case of  $(n = 5, k = 4, d = 4)$  with  $\beta = 48$ . The dashed and the solid lines correspond to the cases of functional and exact repairs, respectively. See (54) and (58) for the corresponding equations. Here, lines 1, 2 and 3 are given by (59), (60) and (61) respectively.

The key idea in the proof lies on the observation that  $\rho(H^{(5)}) \geq \rho(H^{(4)}) \geq \rho(H^{(3)})$ . We will show that (59) and (60) are necessary conditions respectively for  $\rho(H^{(5)}) \geq \rho(H^{(4)})$  and  $\rho(H^{(5)}) \geq \rho(H^{(4)}) \geq \rho(H^{(3)})$  to be true. The following remark underlines an important property of  $\rho(A_{j,j}^{(\ell)})$  preserved in the construction of  $H_j^{(\ell)}$ .

**Remark 1.** The sub-matrices  $A_{j,j}^{(\ell)}, 3 \leq \ell \leq 5, 5 - \ell + 1 \leq j \leq 5$  have full column rank, and  $\rho(A_{j,j}^{(\ell)}) = \rho(H_j^{(\ell)})$ .

$$\begin{array}{c}
 \begin{array}{ccccc}
 H_1^{(5)} & H_2^{(5)} & H_3^{(5)} & H_4^{(5)} & H_5^{(5)} \\
 \left[ \begin{array}{c|c|c|c|c}
 A_{1,1}^{(5)} & A_{1,2}^{(5)} & A_{1,3}^{(5)} & A_{1,4}^{(5)} & A_{1,5}^{(5)} \\
 \hline
 A_{2,1}^{(5)} & A_{2,2}^{(5)} & A_{2,3}^{(5)} & A_{2,4}^{(5)} & A_{2,5}^{(5)} \\
 \hline
 A_{3,1}^{(5)} & A_{3,2}^{(5)} & A_{3,3}^{(5)} & A_{3,4}^{(5)} & A_{3,5}^{(5)} \\
 \hline
 A_{4,1}^{(5)} & A_{4,2}^{(5)} & A_{4,3}^{(5)} & A_{4,4}^{(5)} & A_{4,5}^{(5)} \\
 \hline
 A_{5,1}^{(5)} & A_{5,2}^{(5)} & A_{5,3}^{(5)} & A_{5,4}^{(5)} & A_{5,5}^{(5)}
 \end{array} \right] \\
 H^{(5)}
 \end{array}
 & \rightarrow &
 \begin{array}{ccccc}
 H_2^{(4)} & H_3^{(4)} & H_4^{(4)} & H_5^{(4)} \\
 \left[ \begin{array}{c|c|c|c}
 A_{1,2}^{(4)} & A_{1,3}^{(4)} & A_{1,4}^{(4)} & A_{1,5}^{(4)} \\
 \hline
 A_{2,2}^{(4)} & A_{2,3}^{(4)} & A_{2,4}^{(4)} & A_{2,5}^{(4)} \\
 \hline
 A_{3,2}^{(4)} & A_{3,3}^{(4)} & A_{3,4}^{(4)} & A_{3,5}^{(4)} \\
 \hline
 A_{4,2}^{(4)} & A_{4,3}^{(4)} & A_{4,4}^{(4)} & A_{4,5}^{(4)} \\
 \hline
 A_{5,2}^{(4)} & A_{5,3}^{(4)} & A_{5,4}^{(4)} & A_{5,5}^{(4)}
 \end{array} \right] \\
 H^{(4)}
 \end{array}
 & \rightarrow &
 \begin{array}{ccc}
 H_3^{(3)} & H_4^{(3)} & H_5^{(3)} \\
 \left[ \begin{array}{c|c|c}
 A_{1,3}^{(3)} & A_{1,4}^{(3)} & A_{1,5}^{(3)} \\
 \hline
 A_{2,3}^{(3)} & A_{2,4}^{(3)} & A_{2,5}^{(3)} \\
 \hline
 A_{3,3}^{(3)} & A_{3,4}^{(3)} & A_{3,5}^{(3)} \\
 \hline
 A_{4,3}^{(3)} & A_{4,4}^{(3)} & A_{4,5}^{(3)} \\
 \hline
 A_{5,3}^{(3)} & A_{5,4}^{(3)} & A_{5,5}^{(3)}
 \end{array} \right] \\
 H^{(3)}
 \end{array}
 \end{array}$$

Fig. 12. The matrices  $H^{(5)}, H^{(4)}$  and  $H^{(3)}$ , and the associated block submatrix representations for the case  $n = 5$ . The matrix  $H^{(5)} = H_{repair}$ ,  $H^{(4)}$  is defined based on  $H^{(5)}$ , and  $H^{(3)}$  is defined based on  $H^{(4)}$ .

### B. Proof of (59)

We will be using the rank comparison  $\rho(H^{(5)}) \geq \rho(H^{(4)})$  to prove (59). It follows from (48), (51) and (53) that

$$\rho(H^{(5)}) \geq \rho(A_{1,1}^{(5)}) + \sum_{j=2}^5 \left\{ \left( \rho(A_{j,j}^{(5)}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right)^+ \right\}. \quad (62)$$

We introduce slack variables  $\{\alpha_j, 2 \leq j \leq 5\}$  that take non-negative integer values to convert (51) into equalities i.e.,

$$\delta_j = \left( \rho(A_{j,j}^{(5)}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right)^+ + \alpha_j, \quad 2 \leq j \leq 5. \quad (63)$$

Hence, using (53) we have:

$$\rho(H^{(5)}) = \rho(A_{1,1}^{(5)}) + \sum_{j=2}^5 \left\{ \left( \rho(A_{j,j}^{(5)}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right)^+ + \alpha_j \right\}. \quad (64)$$

$\rho(H^{(4)})$  can be bounded from below quite similar to (62) (see Remark 1 also) to obtain

$$\rho(H^{(4)}) \geq \rho(A_{2,2}^{(4)}) + \sum_{j=3}^5 \left\{ \left( \rho(A_{j,j}^{(4)}) - \sum_{\ell=2}^{j-1} \rho(A_{j,\ell}^{(4)}) \right)^+ \right\}. \quad (65)$$

Our aim at first is to find a lower bound for  $\sum_{j=2}^5 \alpha_j$ . The analysis in Sec. VI-B in fact works with the trivial lower bound  $\sum_{j=2}^5 \alpha_j \geq 0$ . But here, we substitute (64) and (65) in

$$\rho(H^{(5)}) \geq \rho(H^{(4)})$$

to obtain a much tighter lower bound for  $\sum_{j=2}^5 \alpha_j$ . Using this tighter bound in (64), we will obtain a lower bound for  $\rho(H^{(5)})$  in terms of  $\{\rho(A_{i,j}^{(5)}), \rho(A_{i,j}^{(4)})\}$ . We know that the terms  $\{\rho(A_{i,j}^{(5)})\}$  can be expressed in terms of  $\alpha$  and  $\beta$ . In the following Lem. VII.2, we show how  $\{\rho(A_{i,j}^{(4)})\}$  can be expressed in terms of  $\{\rho(A_{i,j}^{(5)})\}$ . Finally, all the terms involve  $\{\rho(A_{i,j}^{(5)})\}$ , and this will lead to the proof of (59).

**Lemma VII.2.** *The following statements hold:*

a)

$$\rho(A_{j,j}^{(4)}) = \rho(A_{j,j}^{(5)}) - \left\{ \left( \rho(A_{j,j}^{(5)}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right)^+ + \alpha_j \right\}, \quad 2 \leq j \leq 5. \quad (66)$$

b)

$$\sum_{\ell=2}^{j-1} \rho(A_{j,\ell}^{(4)}) \leq \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) - \rho(A_{j,j}^{(4)}), \quad 3 \leq j \leq 5. \quad (67)$$

*Proof:* The proof is relegated to Appendix B. ■

By making use of Lem. VII.2, we first obtain a lower bound on  $\sum_{j=2}^5 \alpha_j$ , and subsequently a lower bound on  $\rho(H^{(5)})$  all in terms of  $\{\rho(A_{i,j}^{(5)})\}$ :

$$\sum_{j=2}^5 \alpha_j \geq \frac{1}{3} \left\{ -\rho(A_{1,1}^{(5)}) + \rho(A_{2,2}^{(5)}) + 2 \sum_{j=3}^5 \rho(A_{j,j}^{(5)}) - \left[ 2 \left( \rho(A_{2,2}^{(5)}) - \rho(A_{2,1}^{(5)}) \right)^+ + 3 \sum_{j=3}^5 \left( \rho(A_{j,j}^{(5)}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right)^+ + \sum_{j=3}^5 \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right] \right\} \quad (68)$$

$$\rho(H^{(5)}) \geq \frac{1}{3} \left\{ 2 \sum_{j=1}^5 \rho(A_{j,j}^{(5)}) - \sum_{j=2}^5 \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right\}. \quad (69)$$

Finally, we apply  $\rho(A_{j,j}^{(5)}) = \alpha, 1 \leq j \leq 5$  and  $\rho(A_{i,j}^{(5)}) \leq \beta, 1 \leq i, j \leq 5, i \neq j$  to complete the proof of (59).

### C. Proof of (60)

While proving (59), we leveraged upon the inequality

$$\rho(H^{(5)}) \geq \rho(H^{(4)}).$$

Here, we will make use of the chain

$$\rho(H^{(5)}) \geq \rho(H^{(4)}) \geq \rho(H^{(3)}),$$

to prove (60). First, we consider  $\rho(H^{(4)}) \geq \rho(H^{(3)})$  and obtain a lower bound on  $\rho(H^{(4)})$ . This is carried out precisely the same way as how we obtained the lower bound (69) on  $\rho(H^{(5)})$ . The only change required will be to adapt Lem. VII.2 to express  $\{A_{i,j}^{(4)}\}$  in terms of  $\{A_{i,j}^{(3)}\}$ . Thus we obtain that

$$\rho(H^{(4)}) \geq \frac{1}{3} \left\{ 2 \sum_{j=2}^5 \rho(A_{j,j}^{(4)}) - \sum_{j=3}^5 \sum_{\ell=2}^{j-1} \rho(A_{j,\ell}^{(4)}) \right\}. \quad (70)$$

Observe that (70) is same as (69) except for that  $\{A_{i,j}^{(5)}\}$  are replaced with  $\{A_{i,j}^{(4)}\}$ . The limits of the summation are also modified accordingly.

We next consider the inequality  $\rho(H^{(5)}) \geq \rho(H^{(4)})$  where  $\rho(H^{(4)})$  is lower bounded as in (70) and  $\rho(H^{(5)})$  is equated using (64). It follows that

$$\rho(A_{1,1}^{(5)}) + \sum_{j=2}^5 \left\{ \left( \rho(A_{j,j}^{(5)}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right)^+ + \alpha_j \right\} \geq \frac{1}{3} \left\{ 2 \sum_{j=2}^5 \rho(A_{j,j}^{(4)}) - \sum_{j=3}^5 \sum_{\ell=2}^{j-1} \rho(A_{j,\ell}^{(4)}) \right\}. \quad (71)$$

After invoking Lem. VII.2, we obtain the lower bound:

$$\sum_{j=2}^5 \alpha_j \geq \frac{1}{6} \left\{ -3\rho(A_{1,1}^{(5)}) + 3 \sum_{j=2}^5 \rho(A_{j,j}^{(5)}) - \left[ 6 \sum_{j=2}^5 \left( \rho(A_{j,j}^{(5)}) - \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right)^+ + \sum_{j=2}^5 \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right] \right\}. \quad (72)$$

Substituting (72) back in (64), we obtain the following lower bound on  $\rho(H^{(5)})$ :

$$\rho(H^{(5)}) \geq \frac{1}{6} \left\{ 3 \sum_{j=1}^5 \rho(A_{j,j}^{(5)}) - \sum_{j=2}^5 \sum_{\ell=1}^{j-1} \rho(A_{j,\ell}^{(5)}) \right\}. \quad (73)$$

Finally, we apply  $\rho(A_{j,j}^{(5)}) = \alpha, 1 \leq j \leq 5$  and  $\rho(A_{i,j}^{(5)}) \leq \beta, 1 \leq i, j \leq 5, i \neq j$  on (73) to complete the proof of (60).



### VIII. AN UPPER BOUND ON THE FILE SIZE OF LINEAR ER CODES FOR GENERAL $(n, k = n - 1, d = n - 1)$

In this section, we generalize the result proved for  $(n = 5, k = 4, d = 4)$  in Sec. VII to  $(n, k = n - 1, d = n - 1)$ . We will only provide a sketch of the proofs, as the techniques remain the same as those presented in Sec. VII (see [27] for details). Again, the upper bound on the file size is a direct corollary of a lower bound on  $\text{rank}(H)$  and the bound is achievable using layered codes (see Sec. IX-B). In the following theorem, a lower bound on  $\text{rank}(H)$  is established.

**Theorem VIII.1.** *Consider an ER linear regenerating code  $\mathcal{C}_{lin}$  with full-parameter set  $\{(n, k = n - 1, d = n - 1), (\alpha, \beta)\}$  with  $n \geq 4$ . Let  $H$  denote a parity-check matrix of  $\mathcal{C}_{lin}$ . Then*

$$\text{rank}(H) \geq \begin{cases} \left\lceil \frac{2rn\alpha - n(n-1)\beta}{r^2 + r} \right\rceil, & \frac{d\beta}{r} \leq \alpha \leq \frac{d\beta}{r-1}, \quad 2 \leq r \leq n-2 \\ 2\alpha - \beta, & \frac{d\beta}{n-1} \leq \alpha \leq \frac{d\beta}{n-2} \end{cases}. \quad (74)$$

The corresponding theorem Thm.VII.1 for  $(n = 5, k = 4, d = 4)$  established that  $\text{rank}(H)$  is lower bounded by a piecewise linear curve determined by 3 inequalities. Here, we show such a behavior exists in general i.e.,  $\text{rank}(H)$  can be lower bounded by a piecewise linear curve determined by  $(n - 2)$  inequalities. The last inequality

$$\text{rank}(H) \geq 2\alpha - \beta, \quad (75)$$

is already established by (54), since  $\alpha \geq \beta$  and  $(\alpha - (j - 1)\beta)^+ \geq 0$  for  $3 \leq j \leq n$ . Therefore to complete the proof, it remains to prove the following  $(n - 3)$  bounds on  $\text{rank}(H)$ , ignoring the range of  $\alpha$ :

$$\text{rank}(H) \geq \frac{2rn\alpha - n(n-1)\beta}{r^2 + r}, \quad (76)$$

parameterized by  $2 \leq r \leq n - 2$ . We will set up some notations, and introduce a key lemma that are essential in describing a sketch of the proof.

#### A. Notations and a Key Lemma

1) *The Matrices  $\{H^{(t)}, 3 \leq t \leq n\}$ :* For any matrix  $M$  over  $\mathbb{F}$ , we carry over the notation  $\rho(M)$ ,  $\mathcal{S}(M)$  from Sec. VII-A. Quite similar to the definition of  $H^{(5)}$  in Sec. VII-A, we define  $H^{(n)} = H_{repair}$ , where  $H_{repair}$  is as defined by Lem. VI.2. We denote by  $H_j^{(n)}$  the  $j^{\text{th}}$  thick column of  $H^{(n)}$ ,  $1 \leq j \leq n$ , i.e.,

$$H^{(n)} = [H_1^{(n)} \ H_2^{(n)} \ \dots \ H_n^{(n)}].$$

Next, we define the matrices  $H^{(t)}, 3 \leq t \leq n - 1$  in an iterative manner as follows:

Step 1. Let  $t = n - 1$ .

Step 2. Define the matrices  $H_j^{(t)}, n - t + 1 \leq j \leq n$ , such that the columns of  $H_j^{(t)}$  form a basis for the vector space  $\mathcal{S}(H_j^{(t+1)}) \cap \mathcal{S}(H^{(t+1)}|_{\{n-t, n-t+1, \dots, j-1\}})$ .

Step 3. Define the matrix  $H^{(t)}$  as

$$H^{(t)} = [H_{n-t+1}^{(t)} \ H_{n-t+2}^{(t)} \ \dots \ H_n^{(t)}]. \quad (77)$$

Step 4. If  $t \geq 4$ , decrement  $t$  by 1 and go back to Step 2.

Clearly, the ranks of the matrices  $H^{(t)}, 3 \leq t \leq n$  are ordered as

$$\rho(H^{(t)}) \geq \rho(H^{(t-1)}), \quad 4 \leq t \leq n. \quad (78)$$

We use the notation  $H_j^{(t)}, n - t + 1 \leq j \leq n$  to refer to the  $j^{\text{th}}$  thick column of the matrix  $H^{(t)}$ . While every thick column of  $H^{(n)}$  has exactly  $\alpha$  thin columns, thick columns of  $H^{(t)}, 3 \leq t \leq n - 1$  need not have the same number of thin columns. We point out for clarity that the thick columns of the matrix  $H^{(t)}$  are indexed using  $\{n - t + 1, \dots, n\}$ . We have avoided  $\{1, \dots, t\}$  for the convenience of notation.

2) *Block Matrix Representation of the Matrix  $H^{(t)}$*  : Since  $H^{(n)} = H_{\text{repair}}$ , it has a block matrix representation as given in (45). We write in short-hand

$$H^{(n)} = \left( A_{i,j}^{(n)}, 1 \leq i, j \leq n \right), \quad (79)$$

where  $A_{i,i}^{(n)} = I_\alpha, 1 \leq i \leq n$ . We introduce block matrix representations for  $H^{(t)}, 3 \leq t \leq n-1$  as

$$H^{(t)} = \left( A_{i,j}^{(t)}, 1 \leq i \leq n, n-t+1 \leq j \leq n \right), \quad (80)$$

where  $A_{i,j}^{(t)}$  is an  $\alpha \times \rho(H_j^{(t)})$  matrix over  $\mathbb{F}$  such that

$$\mathcal{S}(A_{i,j}^{(t)}) \subseteq \mathcal{S}(A_{i,j}^{(t+1)}) \cap \sum_{\ell=n-t}^{j-1} \mathcal{S}(A_{i,\ell}^{(t+1)}). \quad (81)$$

Note that (81) is a direct consequence of our definition of the matrix  $H^{(t)}$ . Having set up the notation, we introduce the key lemma that establishes the relations among the ranks of the sub-matrices of  $\{H^{(t)}\}$ . The lemma is similar in spirit to Lem. VII.2, and its proof is omitted here.

**Lemma VIII.2.** a) For any  $t, j$  such that  $3 \leq t \leq n$  and  $n-t+1 \leq j \leq n$ , we have

$$\rho(H_j^{(t)}) = \rho(A_{j,j}^{(t)}). \quad (82)$$

b) For any  $t, j$  such that  $3 \leq t \leq n-1$  and  $n-t+1 \leq j \leq n$ , we have

$$\rho(A_{j,j}^{(t)}) = \rho(A_{j,j}^{(t+1)}) - \left\{ \rho(H^{(t+1)}|_{\{n-t, \dots, j\}}) - \rho(H^{(t+1)}|_{\{n-t, \dots, j-1\}}) \right\}. \quad (83)$$

c) For any  $t, j$  such that  $3 \leq t \leq n-1$  and  $n-t+2 \leq j \leq n$ , we have

$$\rho(A_{j,j}^{(t)}) + \sum_{\ell=n-t+1}^{j-1} \rho(A_{j,\ell}^{(t)}) \leq \sum_{\ell=n-t}^{j-1} \rho(A_{j,\ell}^{(t+1)}). \quad (84)$$

### B. On The Proof of (76)

The bounds in (76) is obtained as a necessary condition for satisfying the chain of inequalities given by

$$\rho(H^{(n)}) \geq \rho(H^{(n-1)}) \geq \dots \geq \rho(H^{(n-r+1)}). \quad (85)$$

In the analysis of (85), we consider in the first step, the inequality  $\rho(H^{(n-r+2)}) \geq \rho(H^{(n-r+1)})$  and obtain a lower bound on  $\rho(H^{(n-r+2)})$ . In the second step, we move on to the inequality  $\rho(H^{(n-r+3)}) \geq \rho(H^{(n-r+2)})$  and obtain a lower bound on  $\rho(H^{(n-r+3)})$ . In the second step, we would make use of a lower bound on  $\rho(H^{(n-r+2)})$  that was derived in the first step. This procedure is continued iteratively until we arrive at lower bound for  $\rho(H^{(n)})$ . The following theorem is a key intermediate step in this process.

**Theorem VIII.3.** For any  $s$  such that  $1 \leq s \leq n-3$ , and any  $t$  such that  $3+s \leq t \leq n$ , the rank of the matrix  $H^{(t)}$  is lower bounded by

$$\rho(H^{(t)}) \geq \frac{2}{(s+1)(s+2)} \left\{ (s+1) \sum_{j=n-t+1}^n \rho(A_{j,j}^{(t)}) - \sum_{j=n-t+2}^n \sum_{\ell=n-t+1}^{j-1} \rho(A_{j,\ell}^{(t)}) \right\}. \quad (86)$$

*Proof:* The proof is by induction on  $s$ , and see [27] for details. ■

One can identify Thm. VIII.3 in the context of  $(n=5, k=4, d=4)$ . The bounds would then be associated with  $(s=1, t=5)$ ,  $(s=1, t=4)$  and  $(s=2, t=5)$ , and are precisely those given in (69), (70) and (73) respectively. To complete the proof of (76), we evaluate the bound in (86) for the  $(n-3)$  pairs given by  $(s, t=n), 1 \leq s \leq n-3$ . By substituting the constraints  $\rho(A_{j,j}^{(n)}) = \alpha, 1 \leq j \leq n$  and  $\rho(A_{i,j}^{(n)}) \leq \beta, 1 \leq i, j \leq n, i \neq j$ , we finally obtain that

$$\rho(H^{(n)}) \geq \frac{2}{(s+1)(s+2)} \left\{ (s+1) \sum_{j=1}^n \alpha - \sum_{j=2}^n (j-1)\beta \right\} = \frac{2(s+1)n\alpha - n(n-1)\beta}{(s+1)(s+2)}, \quad 1 \leq s \leq n-3. \quad (87)$$

By choosing  $r = s+1$ , (76) follows from (87). This completes the proof of (76), and consequently that of the Thm. VIII.1.

## IX. ON THE ACHIEVABILITY OF THE OUTER BOUNDS ON NORMALIZED ER TRADEOFF

The outer bounds presented in the present paper matches with the performance of existing code constructions in certain cases, and we present two such results here.

### A. Characterization of Normalized ER tradeoff for the Case $k = 3, d = n - 1$

In the case of  $k = 3$  and  $d = n - 1$ , the repair-matrix bound is achieved by a construction that appeared in [17]. We will give an example of the repair-matrix bound below:

*Example:* ( $n = 6, k = 3, d = 5$ ) : Using (29), the bound on the ER file size  $B$  can be computed as

$$B \leq \frac{10\alpha}{7} + \frac{34\beta}{7}, \quad 5\beta \geq \alpha > \frac{13\beta}{4}. \quad (88)$$

Based on the bound in (88), an outer bound on the normalized ER tradeoff is drawn in Fig. 4(a). It is required to have a single code construction  $\mathcal{C}_{\text{int}}$  for the normalized operating point  $(\bar{\alpha}_0, \bar{\beta}_0) = (\frac{13}{38}, \frac{2}{19})$  to achieve the entire normalized ER tradeoff, as the remaining points can be achieved by space-sharing of the MSR code  $\mathcal{C}_{\text{MSR}}$ , the MBR code  $\mathcal{C}_{\text{MBR}}$  and the code  $\mathcal{C}_{\text{int}}$ . The construction of  $\mathcal{C}_{\text{int}}$  was provided in [17], and thus establishing that the repair-matrix bound is tight in this case.

### B. Characterization of Normalized ER Tradeoff for the Case $(n, k = n - 1, d = n - 1)$ under the Linear Setting

In the case of linear codes, the bound presented in Theorem I.2 is achieved by canonical layered codes that was introduced in [21]. When specialized to the case of  $d = n - 1$ , the layered codes achieve points described by

$$(\bar{\alpha}, \bar{\beta}) = \left( \frac{r}{n(r-1)}, \frac{r}{n(n-1)} \right), \quad 2 \leq r \leq n-1 \quad (89)$$

on the  $(\bar{\alpha}, \bar{\beta})$ -plane. If one substitutes  $r = 2$  in (89), it corresponds to the MBR point, and the achievable points move closer to the MSR point as  $r$  increases. It is also proved that the point corresponding to  $r = n - 1$  lies on the FR tradeoff in the near-MSR region. An achievable region on the  $(\bar{\alpha}, \bar{\beta})$ -plane is obtained by space-sharing codes for values of  $r$ ,  $2 \leq r \leq n - 1$  along with an MSR-code. We can write the equation of the line segment obtained by connecting two points  $\left( \frac{r}{n(r-1)}, \frac{r}{n(n-1)} \right)$  and  $\left( \frac{(r+1)}{n((r+1)-1)}, \frac{(r+1)}{n(n-1)} \right)$ ,  $2 \leq r \leq n - 2$  as

$$r(r-1)n\bar{\alpha} + n(n-1)\bar{\beta} = r^2 + r, \quad (90)$$

and that of the line segment connecting the MSR point and the point corresponding to  $r = n - 1$  as

$$(n-2)\bar{\alpha} + \bar{\beta} = 1.$$

This matches with the equations of line segments as given in Theorem I.2. The normalized linear tradeoff for  $(n = 6, k = d = 5)$  is given in Fig. 4(b).

## REFERENCES

- [1] B. Sasidharan, K. Senthooor, and P. V. Kumar, "An improved outer bound on the storage-repair-bandwidth tradeoff of exact-repair regenerating codes," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 2430–2434.
- [2] N. Prakash and M. N. Krishnan, "The storage-repair-bandwidth trade-off of exact repair linear regenerating codes for the case  $d = k = n - 1$ ," in *2015 IEEE International Symposium on Information Theory*, June 2015, pp. 859–863.
- [3] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.
- [4] K. W. Shum and Y. Hu, "Cooperative regenerating codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7229–7258, 2013.
- [5] A. M. Kermarrec, N. Le Scouarnec, and G. Straub, "Repairing multiple failures with coordinated and adaptive regenerating codes," in *Proc. IEEE Int. Symp. Network Coding (NetCod)*, Beijing, Jul. 2011, pp. 88–93.
- [6] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [7] Y. Wu, "Existence and construction of capacity-achieving network codes for distributed storage," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 2, pp. 277–288, February 2010.
- [8] K. V. Rashmi, N. B. Shah, and P. V. Kumar, "Optimal Exact-Regenerating Codes for Distributed Storage at the MSR and MBR Points via a Product-Matrix Construction," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5227–5239, Aug. 2011.
- [9] V. Cadambe, S. A. Jafar, H. Maleki, K. Ramchandran, and C. Suh, "Asymptotic interference alignment for optimal repair of mds codes in distributed storage," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2974–2987, 2013.

- [10] D. Papailiopoulos, A. Dimakis, and V. Cadambe, "Repair Optimal Erasure Codes through Hadamard Designs," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3021–3037, 2013.
- [11] C. Suh and K. Ramchandran, "Exact-repair MDS code construction using interference alignment," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1425–1442, 2011.
- [12] N. B. Shah, K. V. Rashmi, P. V. Kumar, and K. Ramchandran, "Interference Alignment in Regenerating Codes for Distributed Storage: Necessity and Code Constructions," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2134–2158, Apr. 2012.
- [13] —, "Distributed Storage Codes With Repair-by-Transfer and Nonachievability of Interior Points on the Storage-Bandwidth Tradeoff," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1837–1852, Mar. 2012.
- [14] I. Tamo, Z. Wang, and J. Bruck, "Zigzag codes: MDS array codes with optimal rebuilding," *IEEE Trans. Inform. Theory*, vol. 59, no. 3, pp. 1597–1616, 2013.
- [15] C. Tian, "Characterizing the rate region of the (4, 3, 3) exact-repair regenerating codes," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 967–975, May 2014.
- [16] "Information theory inequality prover," <http://user-www.ie.cuhk.edu.hk/~ITIP/>, accessed: 2016-Jun-02.
- [17] K. Senthoo, B. Sasidharan, and P. V. Kumar, "Improved layered regenerating codes characterizing the exact-repair storage-repair bandwidth tradeoff for certain parameter sets," in *2015 IEEE Information Theory Workshop, Jerusalem*, April 2015, pp. 1–5.
- [18] I. M. Duursma, "Outer bounds for exact repair codes," *CoRR*, vol. abs/1406.4852, 2014.
- [19] —, "Shortened regenerating codes," *CoRR*, vol. abs/1505.00178, 2015.
- [20] C. Tian, "A note on the rate region of exact-repair regenerating codes," *CoRR*, vol. abs/1503.00011, 2015.
- [21] C. Tian, B. Sasidharan, V. Aggarwal, V. A. Vaishampayan, and P. V. Kumar, "Layered exact-repair regenerating codes via embedded error correction and block designs," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1933–1947, 2015.
- [22] S. Mohajer and R. Tandon, "New bounds on the (n, k, d) storage systems with exact repair," in *2015 IEEE International Symposium on Information Theory*, June 2015, pp. 2056–2060.
- [23] S. Goparaju, S. E. Rouayheb, and R. Calderbank, "New codes and inner bounds for exact repair in distributed storage systems," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 1036–1040.
- [24] M. Elyasi, S. Mohajer, and R. Tandon, "Linear exact repair rate region of (k + 1, k, k) distributed storage systems: A new approach," in *2015 IEEE International Symposium on Information Theory*, June 2015, pp. 2061–2065.
- [25] R. W. Yeung, "A framework for linear information inequalities," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1924–1934, 1997.
- [26] "Solutions of computed information theoretic limits (scitl)," <http://web.eecs.utk.edu/~ctian1/SCITL.html>, accessed: 2016-Apr-17.
- [27] N. Prakash and M. N. Krishnan, "The storage-repair-bandwidth trade-off of exact repair linear regenerating codes for the case d = k = n-1," *CoRR*, vol. abs/1501.03983, 2015.

## APPENDIX A

### PROOF OF THEOREM III.4

Two different estimates on the joint entropy of certain repair data, expressed as functions of  $\{\omega_i\}_{i=0}^{k-1}$ , are used to derive a lower bound on  $\epsilon = \hat{B} - B$ . The repair data considered differ based on the value of  $\mu$ .

*Case 1:*  $\mu \in \{1, 2, \dots, k-2\}$

We set  $r_0 = \left\lfloor \frac{k-\mu}{\mu+1} \right\rfloor$ . We will have two sub-cases for  $r_0 \geq 1$  and  $r_0 = 0$ .

*Case 1(a):*  $r_0 \geq 1$

We consider the sub-trapezoid  $Z_{q,t}$  with parameter  $q = \mu$  and  $t = r_0(\mu+1)$ . Pictorially, it is marked as a trapezium  $EFGH$  in the repair matrix shown in Fig. 13(a). The set of nodes  $T = \{\mu+1, \mu+2, \dots, (r_0+1)(\mu+1)-1\}$  that are repaired by  $Z_{q,t}$  is split into  $r_0$  groups of  $(\mu+1)$  nodes in order, and the corresponding subsets of  $Z_{q,t}$  are denoted by  $\mathcal{E}_i, i = 1, 2, \dots, r_0$ . Pictorially,  $\mathcal{E}_1$  is associated with the trapezium  $EF G_1 H_1$  in Fig. 13(a). Similarly every  $\mathcal{E}_i$  is associated with a smaller trapezium contained within  $EFGH$ . The set  $\mathcal{E}_i$  can again be viewed as the union of two subsets  $\mathcal{V}_i$  and  $\mathcal{T}_i$ , respectively associated with the largest rectangle within the trapezium, and the remaining triangular region. These sets are formally defined as

$$\begin{aligned} \mathcal{E}_i &= \{S_x^y \mid S_x^y \in Z_{q,t}, (\mu+1)i \leq y \leq (\mu+1)(i+1)-1\}, \quad i = 1, 2, \dots, r_0 \\ \mathcal{V}_i &= \{S_x^y \mid S_x^y \in \mathcal{E}_i, (\mu+1)(i+1) \leq x \leq d+1\}, \quad i = 1, 2, \dots, r_0 \\ \mathcal{T}_i &= \{S_x^y \mid S_x^y \in \mathcal{E}_i, (\mu+1)i+1 \leq x \leq (\mu+1)(i+1)-1\}, \quad i = 1, 2, \dots, r_0. \end{aligned}$$

Note that  $\mathcal{E}_i = \mathcal{V}_i \cup \mathcal{T}_i$ . Next, we bound the joint entropy  $H(Z_{q,t})$  as

$$\begin{aligned} H(Z_{q,t}) &\leq \sum_{i=1}^{r_0} H(\mathcal{V}_i) + \sum_{i=1}^{r_0} H(\mathcal{T}_i) \\ &\leq \sum_{i=1}^{r_0} (d - (i+1)(\mu+1) + 2) \cdot [\beta + \mu\theta + \mu\omega_\mu + ((\omega_\mu + \omega_{\mu+1}))] + \sum_{i=1}^{r_0} \frac{(\mu+1)\mu\beta}{2}. \end{aligned} \quad (91)$$

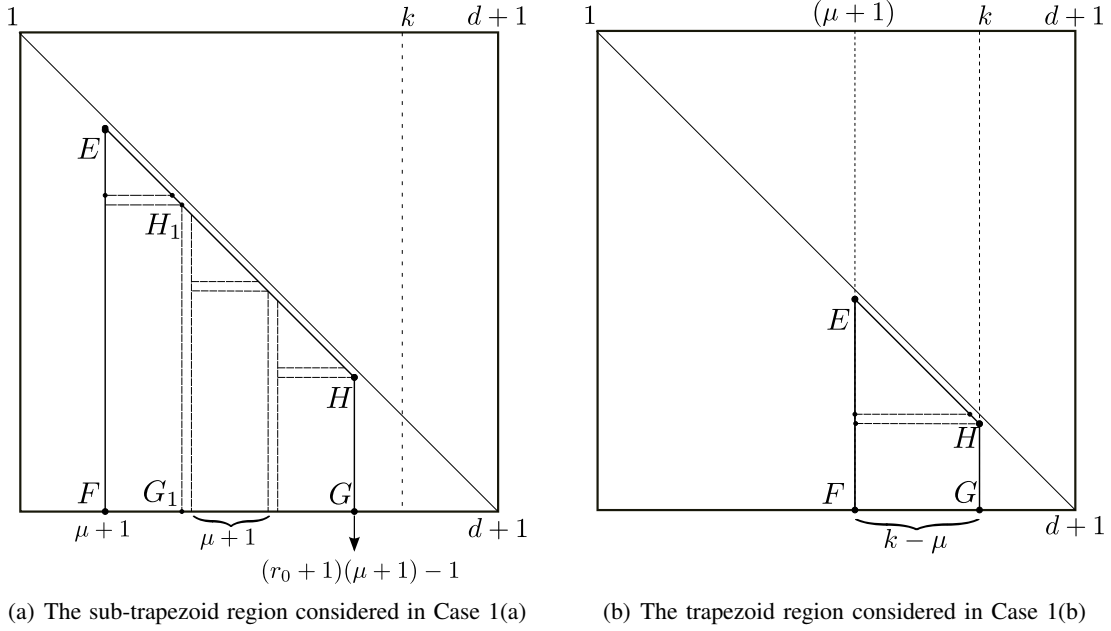


Fig. 13. The illustration of the trapezoid regions considered for Case 1.

In the second inequality, we use (27) of Cor. III.3 to obtain the upper bound on  $H(\mathcal{V}_i)$ . On the other hand, using Lem. III.1, we also have,

$$\begin{aligned}
 H(Z_{q,t}) &\geq H(Z_{q,t} | W_Q) \geq \sum_{i=\mu}^{(r_0+1)(\mu+1)-2} \min\{\alpha, (d-i)\beta\} - \sum_{i=\mu}^{(r_0+1)(\mu+1)-2} \omega_i \\
 &= \left[ \sum_{i=\mu}^{(r_0+1)(\mu+1)-2} (d-i)\beta \right] - \theta - \sum_{i=\mu}^{(r_0+1)(\mu+1)-2} \omega_i.
 \end{aligned} \tag{92}$$

Matching the bounds in (91) and (92) and using the identity (23), we obtain that

$$\epsilon \geq \frac{\left(d - \frac{(\mu+1)(r_0+3)}{2} + 2\right) r_0 \mu (\beta - \theta) - \theta}{\left(d - \frac{(\mu+1)(r_0+3)}{2} + 2\right) r_0 (\mu + 1) + 1}. \tag{93}$$

*Case 1(b):  $r_0 = 0$*

The collection  $Z_q$  of repair data considered in this case corresponds to the trapezoid configuration  $Z_q$  with  $q = \mu$ . The set  $Z_q$  is written as  $Z_q = \mathcal{V} \cup \mathcal{T}$ , where

$$\begin{aligned}
 \mathcal{V} &= \{S_x^y \mid S_x^y \in Z_q, k+1 \leq x \leq d+1\}, \\
 \mathcal{T} &= \{S_x^y \mid S_x^y \in Z_q, \mu+2 \leq x \leq k\}
 \end{aligned}$$

Pictorially,  $Z_q$  is represented by the trapezium  $EFGH$  in Fig. 13(b). Quite similar to the *Case 1(a)*, we invoke Cor. III.3 to bound  $H(Z_q)$  as

$$\begin{aligned}
 H(Z_q) &\geq H(Z_q | W_Q) \leq H(\mathcal{V}) + H(\mathcal{T}) \\
 &\leq (d - k + 1) \cdot [\beta + (k - \mu - 1)\theta + (k - \mu - 1)\omega_\mu + (\omega_\mu + \omega_{\mu+1})] + \\
 &\quad \frac{(k - \mu - 1)(k - \mu)\beta}{2}.
 \end{aligned} \tag{94}$$

On the other hand, using Lem. III.1,

$$\begin{aligned}
 H(Z_q) &\geq \sum_{i=\mu}^{k-1} \min\{\alpha, (d-i)\beta\} - \sum_{i=\mu}^{k-1} \omega_i \\
 &= \left[ \sum_{i=\mu}^{k-1} (d-i)\beta \right] - \theta - \sum_{i=\mu}^{k-1} \omega_i.
 \end{aligned} \tag{95}$$

Matching the bounds in (94) and (95) and using the identity (23), we obtain that

$$\epsilon \geq \frac{(d-k+1)(k-\mu-1)(\beta-\theta) - \theta}{(d-k+1)(k-\mu) + 1} \tag{96}$$

*Case 2:*  $\mu \in \{0, 1, \dots, k-3\}$

We set  $r_1 = \left\lfloor \frac{k-\mu-1}{\mu+2} \right\rfloor$ . We will have two sub-cases for  $r_1 \geq 1$  and  $r_1 = 0$ . In contrast with *Case 1*, we consider a different trapezoid configuration  $(Q, Z_q)$  with  $q = (\mu+1)$  in *Case 2*. It turns out that this change will help in getting a tighter bound in certain regions of  $(\mu, \theta)$ .

*Case 2(a):*  $r_1 \geq 1$

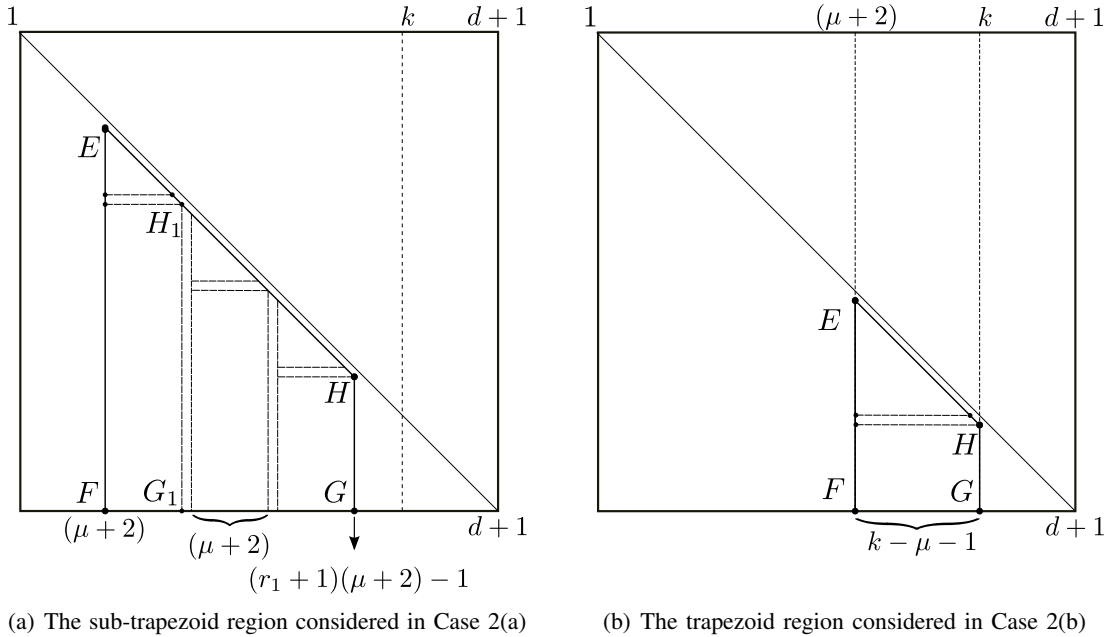


Fig. 14. The illustration of the trapezoid regions considered for Case 2.

In this case, we consider the set  $Z_{q,t}$  with parameter  $q = \mu+1$ ,  $t = r_1(\mu+2)$ . The set of nodes  $T = \{\mu+2, \mu+2, \dots, (r_1+1)(\mu+2)-1\}$  that are repaired by  $Z_{q,t}$  is split into  $r_1$  groups of  $(\mu+2)$  nodes in order, and the corresponding subsets of  $Z_{q,t}$  are denoted by  $\mathcal{E}_i, i = 1, 2, \dots, r_1$ . A pictorial illustration is given in Fig. 14(a). Every  $\mathcal{E}_i$  is further viewed as the union of two subsets  $\mathcal{V}_i$  and  $\mathcal{T}_i$ , respectively associated with the largest rectangle within the trapezium, and the remaining triangular region. The sets of interest are formally defined as

$$\begin{aligned}
 \mathcal{E}_i &= \{S_x^y \mid S_x^y \in Z_{q,t}, (\mu+2)i \leq y \leq (\mu+2)(i+1)-1\}, \quad i = 1, 2, \dots, r_1 \\
 \mathcal{V}_i &= \{S_x^y \mid S_x^y \in \mathcal{E}_i, (\mu+2)(i+1) \leq x \leq d+1\}, \quad i = 1, 2, \dots, r_1 \\
 \mathcal{T}_i &= \{S_x^y \mid S_x^y \in \mathcal{E}_i, (\mu+2)i+1 \leq x \leq (\mu+2)(i+1)-1\}, \quad i = 1, 2, \dots, r_1,
 \end{aligned}$$

where  $\mathcal{E}_i = \mathcal{V}_i \cup \mathcal{T}_i$ . Similar to *Case 1(a)*, we bound the joint entropy  $H(Z_{q,t})$  as

$$H(Z_{q,t}) \leq \sum_{i=1}^{r_1} H(\mathcal{V}_i) + \sum_{i=1}^{r_1} H(\mathcal{T}_i) \quad (97)$$

$$\begin{aligned} &\leq \sum_{i=1}^{r_1} (d - (i+1)(\mu+2) + 2) \cdot [2\beta - \theta + (\mu+1)\omega_{\mu+1} + (\omega_{\mu+1} + \omega_{\mu+2})] + \\ &\quad \sum_{i=1}^{r_1} \frac{(\mu+2)(\mu+1)\beta}{2}. \end{aligned} \quad (98)$$

In the last inequality, we have used (28) of Cor. III.3. On the other hand, using Lem. III.1, we also have,

$$H(Z_{q,t}) \geq H(Z_{q,t} | W_Q) \geq \sum_{i=\mu+1}^{(r_1+1)(\mu+2)-2} \min\{\alpha, (d-i)\beta\} - \sum_{i=\mu+1}^{(r_1+1)(\mu+2)-2} \omega_i \quad (99)$$

$$= \left[ \sum_{i=\mu+1}^{(r_1+1)(\mu+2)-2} (d-i)\beta \right] - \sum_{i=\mu+1}^{(r_1+1)(\mu+2)-2} \omega_i \quad (100)$$

Matching the bounds in (98) and (100) and using the identity (23), we obtain that

$$\epsilon \geq \frac{\left(d - \frac{(\mu+2)(r_1+3)}{2} + 2\right) r_1 [\mu\beta + \theta]}{\left(d - \frac{(\mu+2)(r_1+3)}{2} + 2\right) r_1 (\mu+2) + 1} \quad (101)$$

*Case 2(b):  $r_1 = 0$*

The set  $Z_q$  with  $q = \mu + 1$  is considered in this case. We can write  $Z_q = \mathcal{V} \cup \mathcal{T}$ , where

$$\begin{aligned} \mathcal{V} &= \{S_x^y \mid S_x^y \in Z_q, k+1 \leq x \leq d+1\}, \\ \mathcal{T} &= \{S_x^y \mid S_x^y \in Z_q, \mu+3 \leq x \leq k\}. \end{aligned}$$

A pictorial illustration is given in Fig. 14(b). Following the same line of arguments as in *Case 1(a)*, we obtain that

$$H(Z_q) \leq (d - k + 1) \cdot [2\beta - \theta + (k - \mu - 1)\epsilon] + \frac{(k - \mu - 2)(k - \mu - 1)\beta}{2}, \quad (102)$$

$$H(Z_q) \geq \left[ \sum_{i=\mu+1}^{k-1} (d-i)\beta \right] - \sum_{i=\mu+1}^{k-1} \omega_i. \quad (103)$$

Matching the above two bounds and using the identity (23), we obtain the lower bound for  $\epsilon$ :

$$\epsilon \geq \frac{(d - k + 1) [(k - \mu - 3)\beta + \theta]}{(d - k + 1)(k - \mu - 1) + 1} \quad (104)$$

## APPENDIX B PROOF OF LEM. VII.2

By definition of  $\delta_j$  in (47), we have that

$$\delta_j = \rho(H^{(5)}|_{[j]}) - \rho(H^{(5)}|_{[j-1]}) \quad (105)$$

$$= \dim\left(\mathcal{S}\left(H^{(5)}|_{[j-1]}\right) + \mathcal{S}\left(H_j^{(5)}\right)\right) - \dim\left(\mathcal{S}\left(H^{(5)}|_{[j-1]}\right)\right) \quad (106)$$

$$= \dim\left(\mathcal{S}\left(H_j^{(5)}\right)\right) - \dim\left(\mathcal{S}\left(H^{(5)}|_{[j-1]}\right) \cap \mathcal{S}\left(H_j^{(5)}\right)\right) \quad (107)$$

$$= \rho\left(H_j^{(5)}\right) - \rho\left(H_j^{(4)}\right) \quad (108)$$

$$= \rho\left(A_{j,j}^{(5)}\right) - \rho\left(A_{j,j}^{(4)}\right), \quad (109)$$

where in (107) we used the identity  $\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$  for any two subspaces  $W_1, W_2$ . While (108) follows from the definition of  $H_j^{(4)}$ , (109) from Remark 1. The first assertion (66) of the lemma now follows from (109) and (63).

By definition of  $H_j^{(4)}$ , we have that  $\mathcal{S}(A_{j,j}^{(4)}) \subseteq \sum_{\ell=1}^{j-1} \mathcal{S}(A_{j,\ell}^{(5)})$ , and it follows that

$$\rho(A_{j,j}^{(4)}) \leq \dim\left(\sum_{\ell=1}^{j-1} \mathcal{S}(A_{j,\ell}^{(5)})\right). \quad (110)$$

The RHS of (110) is further upper bounded as follows:

$$\dim\left(\sum_{\ell=1}^{j-1} \mathcal{S}(A_{j,\ell}^{(5)})\right) = \dim\left(\sum_{\ell=1}^{j-2} \mathcal{S}(A_{j,\ell}^{(5)}) + \mathcal{S}(A_{j,j-1}^{(5)})\right) \quad (111)$$

$$\begin{aligned} &= \dim\left(\sum_{\ell=1}^{j-2} \mathcal{S}(A_{j,\ell}^{(5)})\right) + \dim\left(\mathcal{S}(A_{j,j-1}^{(5)})\right) - \\ &\quad \dim\left(\sum_{\ell=1}^{j-2} \mathcal{S}(A_{j,\ell}^{(5)}) \cap \mathcal{S}(A_{j,j-1}^{(5)})\right) \end{aligned} \quad (112)$$

$$\leq \dim\left(\sum_{\ell=1}^{j-2} \mathcal{S}(A_{j,\ell}^{(5)})\right) + \dim\left(\mathcal{S}(A_{j,j-1}^{(5)})\right) - \dim\left(\mathcal{S}(A_{j,j-1}^{(4)})\right) \quad (113)$$

$$= \dim\left(\sum_{\ell=1}^{j-2} \mathcal{S}(A_{j,\ell}^{(5)})\right) + \rho(A_{j,j-1}^{(5)}) - \rho(A_{j,j-1}^{(4)}), \quad (114)$$

where (113) follows from the definition of  $H_j^{(4)}$  and  $A_{j,j-1}^{(4)}$ . If  $j = 3$ , (114) completes the proof of the second assertion. Else, for the case  $j \geq 4$ , the term  $\dim\left(\sum_{\ell=1}^{j-2} \mathcal{S}(A_{j,\ell}^{(4)})\right)$  can further be upper bounded by following a similar sequence of steps as in (111) - (114). This completes the proof.